DOCUMENT RESUME

ED 160 648 TM 007 861

AUTHOR	Gage, N. L.; And Others
TITLE	An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade.
INSTITUTION	Stanford Univ., Calif. Center for Educational Research at Stanford.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	May 78
CONTRACT	NE-C-00-3-0061
GRANT	OB-NIE-G-78-0221
NOTE	196p.; Papers presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978); Tables and parts of document marginally legible due to small type
EDRS PRICE	MF-$0.83 HC-$10.03 Plus Postage.
DESCRIPTORS	*Academic Achievement; Classroom Observation Techniques; Data Analysis; *Effective Teaching; Elementary School Teachers; Grade 3; Inservice Teacher Education; *Parent Participation; Primary Education; *Program Effectiveness; *Student Improvement; Study Skills; Teacher Attitudes; Teacher Behavior; Teacher Characteristics; *Teaching Techniques
IDENTIFIERS	Aptitude Treatment Interaction; Classroom Observation Instruments

ABSTRACT

Correlational findings from four major teacher effectiveness studies were incorporated into a teacher training program whose preliminary results are presented in this set of five papers. N.L. Gage begins by briefly introducing this program's background and rationale. The second paper, Experimental Effects of In-Service Teacher Training Derived from Process-Product Correlations in the Primary Grades, by John Crawford and Jane Stallings, examines teaching skills which are associated with student achievement; instructions for developing teachers' skills are appended. Lyn Corno, in Effects of Parent Instruction on Teacher Structuring and Student Participation in the Third Grade: An Aptitude-Treatment Interaction Approach, discusses the effectiveness of a parent assisted program to develop students' cognitive and behavioral processes which promote successful adjustment to school. Although student attrition and aptitude treatment interaction complicate the analysis of program effects, results are generally positive. In the fourth paper, Nicholas Stayrook and John Crawford present preliminary findings from the analysis of classroom observational data. Finally, Alexis L. Mitman discusses Teacher Self-Report in an Experiment on Teacher Education and Effectiveness. Teachers' reactions to their training and factors affecting teacher training implementation are valuable in interpreting other findings. (Author/GDC)

ED160648

Center for Educational Research at Stanford

BEST COPY AVAILABLE

An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade

Program on Teaching Effectiveness

May 1978

TM007 861

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. L. Gage

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

SCHOOL OF EDUCATION STANFORD UNIVERSITY STANFORD, CALIFORNIA 94305

Stanford University

Center for Educational Research at Stanford

Program on Teaching Effectiveness

May 1978

This monograph contains five papers describing and reporting the presently available results on The Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade. The order of the papers is the same as that in which they were presented at the American Educational Research Association annual meeting in Toronto in March 1978. As of November 1978, a complete report on the experiment, including the results of further analyses not reported here, will be available from ERIC (Educational Resources Information Center).

The papers are listed below:

(1) An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade: Background and Rationale, by N. L. Gage.

(2) Experimental Effects of In-Service Teacher Training Derived from Process-Product Correlations in the Primary Grades, by John Crawford and Jane Stallings.

(3) Effects of Parent Instruction on Teacher Structuring and Student Participation in the Third Grade: An Aptitude-Treatment Interaction Approach, by Lyn Corno.

(4) An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade: The Observational Data, by Nicholas Stayrook and John Crawford.

(5) Teacher Self-Report in an Experiment on Teacher Education and Effectiveness, by Alexis L. Mitman.

The research described in this report was supported by National Institute of Education Contract NE-C-00-3-0061 and by an Interim Grant OB-NIE-G-78-0221 to the Center for Educational Research at Stanford, Program on Teaching Effectiveness. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that office should be inferred.

An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade: Background and Rationale

N. L. Gage

Stanford University

School of Education

Center for Educational Research at Stanford

Program on Teaching Effectiveness

Abstract

This paper provides the historical background and theoretical and methodological rationale for an experiment on teacher training, teacher behavior, and parent-assisted instruction in the third grade. The historical background refers to the recent evolution of research on teaching that has created the need for true experiments. The theoretical and methodological rationale rests on the behavioral implications of process-product correlations and the superior basis for causal inferences provided by true experiments where implementation of treatments is carefully appraised.

An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade: Background and Rationale

This is an introduction to a set of four papers describing and reporting the results of an experiment on teacher training, teacher effectiveness, and parent-assisted instruction in the third grade. The background of the experiment consists of the substantial body of correlational process-product research on teacher effectiveness conducted during the past decade. That research culminated in large measure in four major studies (Stallings & Kaskowitz, 1974; Brophy & Evertson, 1974; McDonald & Elias, 1976; Soar, 1973) that represent, in the aggregate, an investment of several million dollars in the search for teaching behavior correlates of academic achievement and affective outcomes in the primary grades.

The findings of these four studies may be regarded as providing a capstone of the correlational approach to meeting the need for an improved scientific basis for the art of teaching and the education of teachers. A logical next step is to exploit those findings by determining whether the teaching practices implied by the findings of those studies can be fostered through teacher training and whether those teaching practices do in fact causally influence, rather than merely correlate with, pupil achievement.

The work described in this set of papers attempted to take that next step. It converted the correlational findings into recommendations for teaching practice. These recommendations were, in turn, incorporated into a teacher training program. The training program was used in an extremely inexpensive, reading-and-testing, or "minimal," format and also in a more traditional, much more expensive, interactive workshop, or "maximal," format. The two kinds of training and a no-training (observed-only) treatment were given to three groups of third-grade teachers in each of two school districts. The teachers, after being

stratified in each district on the basis of the mean pretested reading achievement of their pupils, were randomly assigned to the three treatments.

Crossed with the three training levels was the use or non-use of a parent-assisted learning skills program intended to help pupils understand the concepts of subject matter organization and participation in academic activities--concepts considered important in facilitating pupil achievement.

One major shortcoming of much previous experimentation on teaching has been inadequate observation to determine the degree to which training treatments were implemented, i.e., were actually reflected in the classroom behavior of teachers and pupils. Accordingly, the present experiment invested heavily in detailed, frequent, specifically relevant observation by carefully trained staff members using a special adaptation of the SRI Classroom Observation Instrument that had been employed in the evaluation of the Planned Variations Program of Project Follow-Through.

Finally, the teachers and their pupils responded to inventories, questionnaires, and tests at the beginning and end of the school year. The teachers were also interviewed, and their classes were described in an ethnographic style by an experienced classroom teacher and research specialist.

All of this effort and the resulting data make possible an appraisal of one approach to the improvement of teacher education, teaching practices, pupil aptitudes, and pupil achievement. In the papers that constitute this set, reports on various facets of the study will be made.

References

Brophy, J., & Evertson, C. Process-product correlations in the Texas Teacher Effectiveness Study: Final Report. Research Report No. 74-4. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1974.

McDonald, R., & Elias, P. The effects of teaching performance on pupil learning, Beginning Teacher Evaluation Study: Phase II, Final Report: Vol. I. Princeton, N.J.: Educational Testing Service, 1976.

Soar, F. Final report, Follow Through classroom process measurement and pupil growth (1970-1971). Gainesville, Fla.: Institute for the Development of Human Resources, University of Florida, 1973.

Stallings, J., & Kaskowitz, D. Follow Through classroom observation evaluation, 1972-1973. Menlo Park, Calif.: Stanford Research Institute, 1974. Office of Education Contract OEC-0-8522480-4633(100).

Experimental Effects of In-Service Teacher Training Derived from Process-Product Correlations in the Primary Grades

John Crawford

Center for Educational Research at Stanford

Jane Stallings

SRI International

Abstract

Two forms of in-service teacher training were compared with no intervention. Analyses of the end-of-year achievement data revealed complex interactions between the teacher training intervention, the parent-assisted Learning Skills Program intervention, and school district membership. Results indicated that other influential factors included teacher verbal ability and structuredness of teaching style. Implications and qualifications of these results are discussed, with reference to the importance of teacher attitudes and environmental considerations.

Experimental Effects of In-Service Teacher Training Derived from Process-Product Correlations in the Primary Grades

Earlier field-based studies of teaching effectiveness have typically yielded descriptive information, or coefficients of correlation between classroom process variables and pupil outcomes, or comparisons of global patterns of teaching style (e.g., "open" vs. "traditional") using quasi-experimental designs such as matched groups. The present investigation was designed to allow statements concerning the cause-and-effect relationships between specific instructional strategies and pupil achievement. The investigation took the form of an experiment employing random assignment of teachers to experimental and control groups and systematically intervening with a teacher training program based on the correlational findings from earlier studies of primary-grade classes.

The results of four large-scale correlational studies were examined in detail to derive the training content. The studies were those by Brophy and Evertson (1974), McDonald and Elias (1976), Stallings and Kaskowitz (1974), and Soar (1973). One of the chapters in Volume I of the final report on the experiment (PTE, 1977), presents the teacher training packets (the materials used in the intervention) in their entirety. The bulk of the treatment consisted of five packets, averaging 10 to 12 pages each. In the margins of the materials as presented in Volume I, codes are given for cross-referencing specific teacher training recommendations to "information sheets" that tell which of the four studies the recommendation is based on, the correlation of a given teaching behavior or process variable with reading achievement; the means, standard deviations and metrics of process variables; and summaries or quotations of the original author's interpretation. These information sheets were prepared to document the information used as the basis for the recommendations to teachers.

The process of moving from research-oriented technical reports to training materials suitable for distribution to practicing teachers necessarily entailed some reduction and synthesis on our part. The initial inspection of the four large reports revealed literally thousands of process-product correlations in the primary grades. The first step was to focus on meaningful (statistically significant or near-significant) correlations with reading achievement in the second and third grades only. This focus still yielded 125 variables spread across the four studies. Not all of these variables were actually used in the teacher training program; the 81 variables that were used as the basis for the training packets were usually those agreeing with theoretical expectations or with other studies. A number of variables were not used because of our inability to operationalize the teaching behavior (measured by the process variable) in a manner that would make it useful for teachers. Of the 81 variables referenced in the training packets, 50% are from Brophy and Evertson (1974), 31% are from Stallings and Kaskowitz (1974), 15% are from McDonald and Elias (1976), and 4% come from Soar (1973).

From this information it can be seen that over 80% of the training prescriptions are directly tied to the findings of either Brophy and Evertson or Stallings and Kaskowitz. The other two studies are "underrepresented" primarily because of their use of composite (factor) scores as process measures--which are typically difficult to operationalize for teachers. This reliance on composite scores was used to a large degree in the McDonald and Elias study and was the only method of analysis in the Soar investigation. While these two studies were not relied on to a great extent in forming recommendations for teachers, it was still possible to establish some agreement between their interpretations and conclusions and the gist of the teacher training program.

It was necessary to develop a conceptual framework inclusive enough to encompass the findings from the four studies while still consisting of categories useful in a field-based experiment with regular classroom teachers. Table 1 gives the major headings of that framework and shows the number of variables from the studies that made up each category:

Table 1

Summary of Framework Encompassing the Correlational Findings

Study[1]	Number of Variables	Category Name
SK ME S	5 8 6	A. Instructional Methods Subcategories: Pupil Self-Instruction Focused Instruction from the Teacher Specific Instructional Techniques
BE	3	B. Classroom Rules
SK BE	5 2	C. Grouping and Instructional Organization Subcategories: Use of Large-group, Small group, or Individual Instruction Instructional Organization
BE SK ME S	3 3 2 2	D. Discipline and Management Subcategories: Pupil Negative Affect Teacher Reaction to Disciplinary Problems Teacher Monitoring, Movement, and Checking

Table 1 (cont'd)

Study	Number of Variables	Category Name
BE SK ME	4 6 3	E. Instructional Materials
BE SK ME S	15 16 2 1	F. Teacher Questioning and Student Responding Subcategories: Teacher Academic Instruction Pupil Academic Responses Non-Academic Interactions Non-Academic Pupil Responses, Questions, and Comments Quality (Correctness) of Pupil Responses Selection of Pupils for Questions
BE SK ME S	35 1 2 1	G. Teacher Feedback Subcategories: Teacher Praise Teacher Criticism Teacher Acceptance of Pupil Comments and Opinions Teacher Feedback that Terminates the Interaction Teacher Feedback that Sustains the Interaction Brief Feedback Extended Feedback

[1] SK = Stallings & Kaskowitz
BE = Brophy & Evertson
ME = McDonald & Elias
S = Soar

At the level of "subcategories" in the above conceptualization, more than 60% were based on findings from more than one study. Only one of the seven major categories ("B. Classroom Rules") was not represented by results from more than one of the studies. This framework was reorganized, for presentation of recommendations to teachers, into three areas:

(1) Behavior Management--consisting of categories B and D in Table 1,

(2) General Instructional Methods--consisting of categories A, C, and E in Table 1,

(3) Questioning and Feedback Techniques--consisting of categories F and G in Table 1.

The first packet contained an introduction and vignette. The recommendations for the three areas were presented to the teachers in the second, third, and fourth packets, respectively. The fifth packet contained a summary and review. Appendix A of this paper presents the fifth training packet.

Before leaving this matter of the connection between the earlier correlational studies and our training intervention, we should discuss the SES-differentiated findings of the Brophy and Evertson study. When they divided their sample of classes into lower and higher SES groups (lower to lower-middle class vs. middle-middle to upper-middle class), they found that the strength and even the direction of the process-product correlations could vary according to the classroom's SES level. Since these results pertained only to one SES-type of student, it was inferred that the training recommendations should specify either that these teaching behaviors are hypothetically "optimal" for only a particular type of student, or that the findings should not be discussed at all.

Since the SES-differentiated findings comprised a large proportion of the significant results from the Brophy and Evertson study (and hence, a large proportion of the

Questioning and Feedback training), a decision was made to develop behavioral characterizations of lower and higher SES pupils. Then these pupil characterizations would be defined for teachers receiving training materials, and the packets would instruct teachers to teach in different fashions to the two different types of pupils. These two pupil types were called "more academically oriented" and "less academically oriented." The behavioral characteristics of these pupils were obtained by careful examination of a technical report (Brophy, Evertson, Crawford, & Sherman, 1975) which presented and discussed pupil behavior differences according to classroom SES-level for the low-inference process variables. These differences were compiled and then integrated into behavioral descriptions of lower- and upper-SES pupils. Then, whenever teachers were given recommendations which pertained to only one type of pupil, they were advised that the particular teaching method was shown to be effective for "more academically oriented" or "less academically oriented" pupils. Analyses that address the teachers' implementation of the academically oriented-differentiated strategies are presented in the Stayrook and Crawford (1978) paper.

While this kind of recommendation did require a substantial extrapolation from the original Brophy and Evertson results (to go from learner attribute data based on a school-level classification of SES to a pupil-level classification on behavioral characteristics), it was considered to be better to use these data and make such an extrapolation than not to consider the findings at all. Although there was no mention of achievement in the behavioral descriptions of less and more "academically oriented" children, the interview data discussed by Mitman (1978) suggest that teachers may have relied heavily on their students' academic achievement in their decision-making about "academic orientation." As can be seen from the section quoted below from the first packet, the construct of academic orientation is probably closer to achievement motivation than to achievement per se:

> . . . One type of student is less academically oriented, frequently answers incorrectly, and is often somewhat unruly. He or she is not very inclined to respond, and teachers often find it necessary to encourage the child to answer questions (by rephrasing, giving clues, etc.). While this type of pupil is generally less well-behaved, teachers do find that interactions of a personal nature help in maintaining rapport and keeping these pupils interested in school.

and

> . . . The other type of student is usually more academically oriented, takes part and answers questions correctly during question-and-answer sessions, and actively seeks out the teacher for help with seatwork assignments. Since these pupils are more interested in the subject matter being learned, teachers do not find it necessary to rely on personal discussion as a means of getting the pupils involved in the learning activities of the classroom. Classrooms composed of these pupils are characterized by more questions from the teacher (and, hence, more responses from pupils) than classrooms composed mostly of pupils of the first type.
>
> (from Teacher Training Packet 1, pages 5 and 6)

It should also be noted that, to the greatest extent possible, the training recommendations were accompanied by discussion of their rationale and the broader concepts behind the specific recommendations.

Since random assignment of teachers to treatments was used in conjunction with an experimenter-developed intervention, the results bear on the causal efficacy of many research-based prescriptions for teaching reading. While multivariate methods make possible the assessment of implementation of the specific and discrete components of the training program, the most persuasive inferential statistics should be those associated with the evaluation of the program as a whole. This proposition follows from the fact that the discrete teaching behaviors were not manipulated independently of each other.

Method

Teachers were assigned to treatment groups by a stratified random assignment procedure within each of the two participating school districts. The experiment had a 3 x 2 factorial design replicated in the two districts (18 classes in one district

and 15 in the other). There were three levels for the teacher training factor and two levels for the Learning Skills Program treatment (a parent-assisted instructional program manipulated independently of the Teacher Training Program and discussed by Corno, 1978). The three groups in the teaching training factor were (a) a control group (observed-only, no intervention), (b) a reading-and-testing ("minimal") training group, and (c) an interactive ("maximal") in-service training group. Groups b and c differed in the mode of in-service training. That is, the content of the training given to groups b and c was intended to be the same; only the delivery system differed. The "minimal" training group received printed materials and self-administered tests on the teaching methods, while the "maximal" training group received the same printed packets and tests and also participated in meetings with PTE staff and other teachers to discuss, practice, and study the techniques, engage in role-playing exercises, and examine videotapes of a "model" teacher performing the behaviors.

The comparison of "minimal" and "maximal" in the design and analysis of the experiment was considered important for at least two major reasons:

(1) This aspect of the experiment permits a kind of "cost effectiveness" perspective on teacher training programs. For example, if the "minimal" training were found to be as effective in some respects as the "maximal" training, school districts with limited funds for in-service training would have an extremely inexpensive method for training teachers to adopt strategies expected to facilitate reading achievement.

(2) One other study, by Anderson and Brophy (1976), has reported evidence that student achievement can be significantly increased by giving teachers printed matter containing recommendations about teaching effectiveness--again, with achievement scores serving as the measure of effectiveness. Although their

data are still being analyzed, the findings reported thus far are promising, and there seems to be little doubt about the value of attempting to confirm their experiment's positive effect on student achievement. Since the content of their training differed from ours, the present study is primarily a replication of the method of training teachers by only giving them materials to read and tests to take. Given the possibility that this might not be sufficiently strong treatment, we also included the more standard, workshop-type training ("maximal" training) as one level of the teacher training factor in this experiment. As will be noted below, the mode of training does not appear to be the sole determinant of many teacher training outcomes.

The study was conducted in three phases:

(1) From mid-September through mid-October 1977, "baseline" low inference observations were conducted one full day each week in all classes. Pretesting of pupils occurred during this period, and teachers filled out a questionnaire dealing with their attitudes, aptitudes, and teaching styles. Classes were assigned to treatments around the end of October.

(2) The training program was conducted during the six-week period beginning in the first week of November (the same period in which the Learning Skills Program was administered). Observation continued on a full day per week schedule during this phase. The "minimal" training group received one packet and one set of related (criterion-referenced) questions per week. The interactive, "maximal" training group got the same materials and came to one two-hour meeting with staff members each week.

(3) From January through mid-May, observations were made about one full day every two weeks with the low-inference system and also with a set of high inference rating scales developed for use after each morning and each afternoon of observation. Test batteries were administered twice during this

posttreatment phase: once in late January, when reading comprehension and understanding of the Learning Skills Program concepts were measured, and again in late May. The May posttesting included measures of all the affective variables (self-concept, attitude toward school, anxiety, and locus of control) and the cognitive outcomes (reading comprehension, vocabulary, and nonverbal reasoning) used in the pretest battery. The training activities occurring between January and May included mailing a "refresher" training packet to both the "minimally" and "maximally" trained groups and, for the maximally trained group, following the "refresher" packet with a videotaping of the teacher and feedback concerning treatment implementation. This videotaping and feedback in the Spring with the maximal group was intended to maintain the same sort of distinction between minimal and maximal training that occurred during the main part of the training--from early November through mid-December.

The full-day observations made frequently between September and May provide unusually rich information concerning classroom activities. These data permit determination of the agreement between our results and the earlier studies as well as the degree to which the training resulted in differences in the observed teaching behaviors. (See Stayrook and Crawford (1978) and Program on Teaching Effectiveness (1978) for a more complete description of this part of the study.)

Results and Discussion

Before summarizing the results of the many statistical analyses performed with the data, we will briefly discuss here the basis of what might be termed a model of research on teaching and learning in schools. It is hoped that this will prove useful, not because the conceptualization explicates heretofore unrecognized important linkages (it probably doesn't), but because it will provide a context for answering the question: "What were the observable effects of the teacher training program?"

First, we should note that this experiment was conducted in regular classes throughout the school year and that end-of-year achievement is the major student outcome of interest. Accordingly, the "sequence-of-effects" model in Figure 1 below deals with relatively "molar" phenomena. The components of the model involve macro-phenomena that generally have social and educational significance.

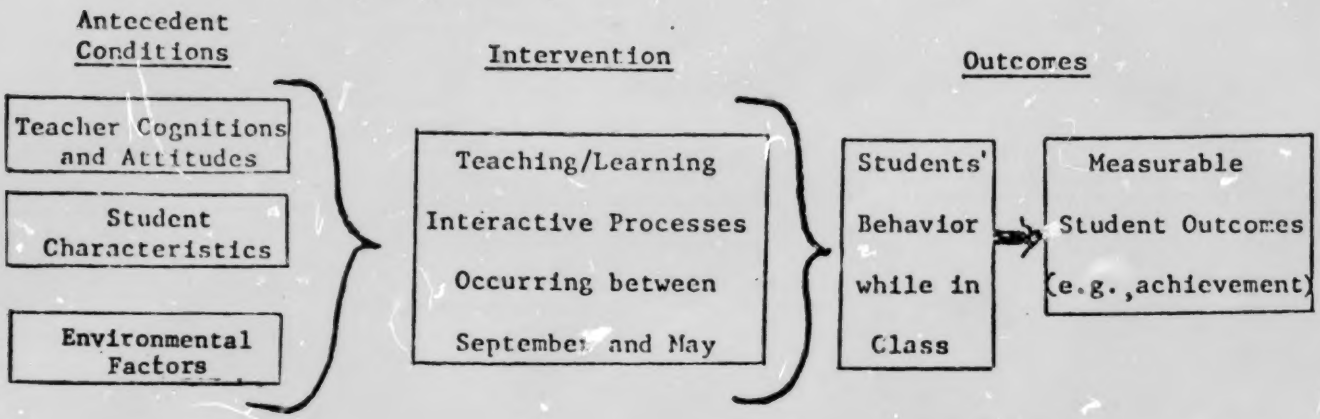


Figure 1. Generalized Sequence-of-Effects Model of Teaching and Learning in Schools

In this model, the "antecedent conditions" relevant to the teaching-learning process are categorized as (a) teacher characteristics (cognitions and attitudes), (b) student characteristics, and (c) environmental factors. In the present study, as shown in Figure 2, two teacher characteristics of apparent import are the teachers' verbal ability and the teacher-reported "structuredness" of teaching style at the beginning of the year. As measured here, structuredness refers to teacher initiation of interaction, emphasis on subject matter, following of lesson plans, seating arrangements (assigned vs. unassigned), etc. The most critical student characteristic (when reading achievement in May is the dependent variable) is the students' achievement level at the beginning of the year. The major environmental factor in this experiment is school district membership; a number of main effects and interactions associated with the district factor will be discussed later in this report.

In the second phase of the model, "intervention," the major concern is the teacher training factor in this experiment. The general term "intervention" is intended to connote a broader meaning than its strictly experimental usage; for example, in a purely correlational study, the teachers still implement an intervention with their students--the intervention being whatever teaching methods and curricula are followed. In this study, the intervention includes not only the set of research-based prescriptions in the Behavior Management, Instructional Methods, and Questioning and Feedback areas, but also other teaching practices in those areas and other teaching methods in entirely different areas. For obvious reasons, we focus on the training-related areas.

The third phase of the model, labeled "outcomes," has two subcomponents: observed student behavior in the classroom and the achievement variables of interest. In this study, the hypotheses were that teachers who followed the training recommendations would find their students less disruptive, exhibiting more time-on-task, and more involved in learning activities. Our results will demonstrate how class attentiveness and passivity (the opposite of "involvement" mentioned above) are determined by treatment group membership.

At the last stage, the students' classroom behavior is assumed to be related to their performance on the dependent variables of interest to the investigator. In this experiment, these achievement variables included reading comprehension and vocabulary. In a test of an affective education program, it might include self-esteem as a dependent variable.

In its general form, as represented in Figure 1, this conceptualization is more methodological than substantive. Any heuristic value of the model will likely result from its pointing to the many factors influencing (sometimes in one direction, sometimes recursively) the teaching-learning process. The message to researchers is that it would be desirable to obtain data on each of

the components in the model; otherwise, potential mediating or confounding effects will not be detected.

This generalized view of teaching and learning in schools takes on somewhat more substance as we insert the relevant constructs from this experiment into the model as specific components:

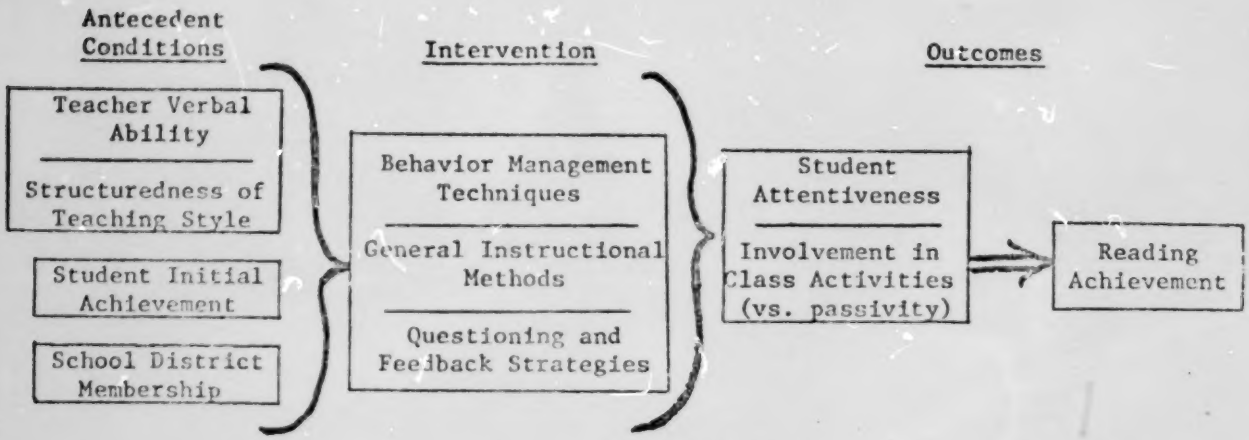


Figure 2. Sequence-of-Effects Model of Teaching and Learning in Schools for the PTE Experiment

This model is amenable to the regression procedure referred to as path analysis, and path analyses will be reported on other occasions. (The causal connections between components would have to be further specified.) In this paper, we will not be reporting path coefficients since we wish to obtain the maximum amount of information about effects and interrelationships. Path analysis assesses the relative strengths of multiple factors linked in a hypothesized causal ordering. At this point it seems more important to piece together a picture from several correlational and inferential analyses. The evidence on the factors represented in Figure 2 is as follows.

Antecedent Conditions

School district membership. As reported in Volume I of the final report, district effects on a number of pretest measures were significant. Most pertinent are the main effects on initial achievement measures ($p < .05$) and nonverbal reasoning ($p < .05$). The direction of the means in each case favored District I over District II. However, when a dummy variable representing District membership (a binary, 1-0 vector) was entered as a covariate along with pretest class mean achievement measures in an analysis of covariance of end-of-year achievement, the nonsignificant F-to-enter for the district variable indicated that most of the observed district differences were due to variation in mean pretest class achievement, and not vice versa. There was also a significant ($p < .05$) district-by-teacher training interaction on adjusted posttest vocabulary, such that District II means showed a strong decreasing trend from the maximal group to the minimal group to the control group, while District I means were in the pattern of an inverted "V"--with that of the minimal training group higher than those of the maximal and the control groups.

Student Characteristics. With the class as the unit of analysis, the pretest and posttest composite "Total Reading" measure (sum of Vocabulary and Reading Comprehension z-scores) correlated .96. This suggests two conclusions: the reliability of the raw scores is very high; and, since over 92% of the variation in class achievement at the end of the year is attributable to achievement in October, there seems to be relatively little "room" for a sizeable treatment effect. (But it would be unwarranted to conclude that the treatment effect cannot therefore be educationally and socially significant. The question of that kind of significance needs to be discussed more fully in another context.)

Teacher Cognitions and Attitudes. Interesting relations were found between the teacher pretest measures and the composite "total implementation" score (based on training and posttraining data). The correlations were .45 ($p < .01$) between verbal ability and posttraining implementation and .28 ($p = .06$) between self-reported structuredness of the teacher's style-and-classroom environment and the implementation measure. The verbal ability measure was adapted from Ryans's questionnaire (Ryans, 1970) and the structuredness instrument was the one used by Stallings and Kaskowitz (1974). Both were administered as part of the questionnaire given to teachers in October. The fact that the means on both of these predictive measures (verbal ability and structuredness) were higher for the minimal group than the maximal group in both districts is undoubtedly important for the interpretation of mean treatment-group differences on total implementation.

Intervention Effects

Composite implementation variables in the form of z-scores were created by combining measures of components of the teacher training program that were not differentiated according to the students' academic orientation. The "total implementation" measure was comprised of six Behavior Management categories, six Instructional Methods components, and four Questioning and Feedback Strategies; definitions of the individual components are given in the Stayrook & Crawford (1978) paper.

The summary scores representing implementation during training and posttraining (12 full days of observational data between November and May) were analyzed with a 3 by 2 analysis of variance (pooling any Learning Skills Program-related variance into the error term). This ANOVA showed no District main effect ($F < 1.0$) and no District by treatment interaction ($F < 1.0$). The training main effect was:

F (2,27) = 4.811, p = .016, with means favoring the minimal training group, then the maximal group and then the control. Analysis conducted on the implementation scores for the three content areas of the training indicated that composite measures in each category showed main effects favoring the training groups (the control group exhibited the least conformity to the set of treatment behaviors): for Behavior Management, F (2,27) = 2.379, p = .112; for Instructional Methods, F (2,27) = 2.763, p = .081; and for Questioning and Feedback, F (2,27) = 2.356, p = .114. Table 2 below gives means and standard deviations of the Z-score implementation variables.

Table 2

Means and Standard Deviations of Treatment Groups for Composite Implementation Variables

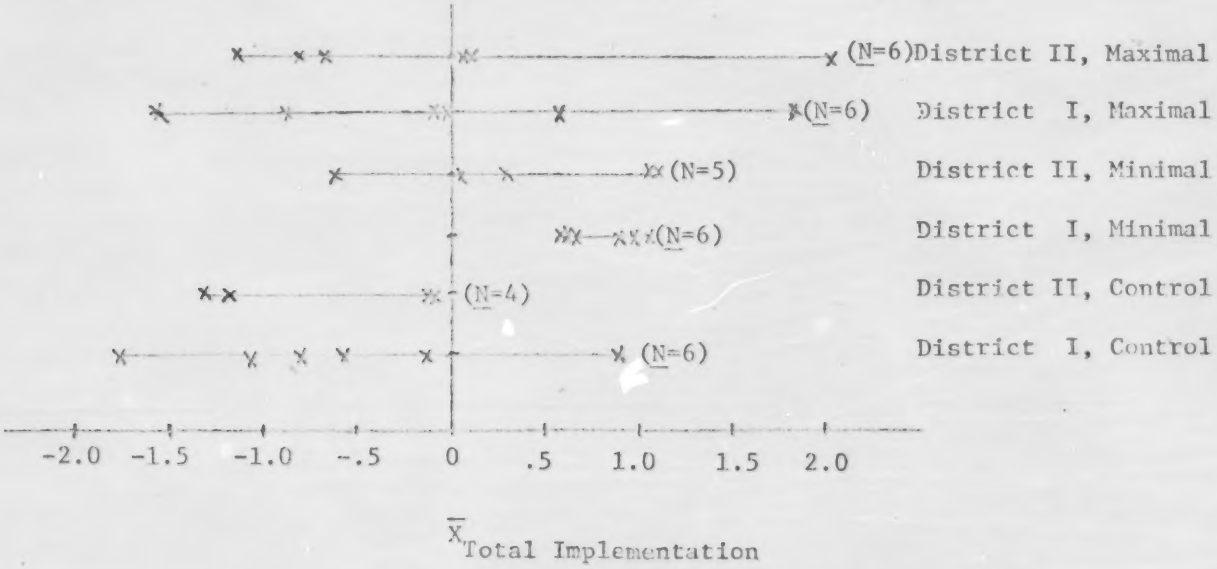
(z-scores with $\bar{X}_{Total}$ = 0.0 and S.D.$_{Total}$ = 1.0)

		Total Implementation		Behavior Management	
Treatment	N	$\bar{X}$	S.D.	$\bar{X}$	S.D.
Maximal	12	-.0565	1.1804	.0035	1.2072
Minimal	11	.6260	.5198	.4406	.5046
Control	10	-.6209	.8005	-.4889	.9942
		Instructional Methods		Questioning and Feedback	
Treatment	N	$\bar{X}$	S.D.	$\bar{X}$	S.D.
Maximal	12	-.1264	1.2626	.0372	.7441
Minimal	11	.5395	.7334	.4245	1.1439
Control	10	-.4418	.6446	-.5115	.9551

The consistently superior implementation by the minimal training group was unexpected. This result may be due in part to the covariation of implementation with certain teacher characteristics, as noted above. By chance, the mean pre-test scores of the minimal group on the teacher characteristics were somewhat

higher than the maximal group means. The higher implementation means for the minimal training group may be, at least partially, an artifact of group differences on the antecedent teacher characteristics. In any case, it is noteworthy that the maximal and minimal training groups combined are about .9 of a standard deviation above the control group on the summary total implementation scores (a "large" effect size).

The graph below shows the individual teacher's scores (by group) for the posttraining implementation score:


(N=6) District II, Maximal
(N=6) District I, Maximal
(N=5) District II, Minimal
(N=6) District I, Minimal
(N=4) District II, Control
(N=6) District I, Control
-2.0 -1.5 -1.0 -.5 0 .5 1.0 1.5 2.0
X Total Implementation

These data indicate that 6 of the 10 control group teachers were more than 1/2 S.D. below the grand mean, while 1 of the 11 minimal teachers and 5 of 12 maximal training teachers had scores below -.5.

If variance in the teacher characteristics variables is partially responsible for the higher implementation of the minimal training group, then the differences between minimal and maximal groups should decrease when a covariance-type adjustment is computed for the implementation measure. The adjusted means in Table 3 below indicate support for this hypothesis.

Table 3

Means and Minimal vs. Maximal Differences on Total Implementation

(a) Unadjusted Means:

$\bar{X}$ minimal = .63
$\bar{X}$ maximal = -.06
$\bar{X}$ control = -.62

Difference between Minimal and Maximal = .69 (standard scores)

(b) Adjusted for Verbal Ability:

$\bar{X}$ minimal = .47
$\bar{X}$ maximal = -.02
$\bar{X}$ control = -.49

Difference between Minimal and Maximal = .49 (standard scores)

(c) Adjusted for Initial Structuredness of Teaching:

$\bar{X}$ minimal = .56
$\bar{X}$ maximal = .01
$\bar{X}$ control = -.62

Difference between Minimal and Maximal = .55 (standard scores)

(d) Adjusted for both Verbal Ability and Structuredness:

$\bar{X}$ minimal = .43
$\bar{X}$ maximal = .00
$\bar{X}$ control = -.48

Difference between Minimal and Maximal = .43 (standard scores)

The pattern of these means shows that the difference between maximal and minimal groups is reduced to about 62% of the unadjusted mean difference--when teacher verbal ability and initial structuredness are taken into account (a reduction of more than one-third). Also, the training main effect on implementation is not significant (p = .163) in the analysis of covariance associated with (d) in Table 3.

Although correlations between the composite implementation variables and residualized achievement were not exceptionally strong, they do present a consistent pattern of results. When the implementation scores for the three content areas and the total implementation measure were used as process indices in a process-outcome matrix with residualized vocabulary, comprehension, and total reading, 9 out of 12 rs were positive. The measure of Behavior Management implementation correlated .21 ($p = .12$) with residualized vocabulary, -.08 with reading comprehension and .10 with residualized total reading. The summary score for Instructional Methods also correlated slightly positively with residualized vocabulary and total reading (non-significantly) and slightly negatively with residualized reading comprehension ($r = -.003$). Questioning and Feedback implementation was strongly correlated with vocabulary and total reading (rs of .37 and .35, $p < .05$, respectively for vocabulary and reading), and slightly positively with reading comprehension ($r = .13$). The total implementation score related positively to residual vocabulary and total reading (rs of .28 and .20, respectively), but was not correlated ($r = -.02$) with reading comprehension. In general, these findings suggest stronger influences of implementation on vocabulary than on reading comprehension (the rs with total reading seem primarily due to the vocabulary component).

The training materials and the workshop meetings held out to teachers the likelihood that, if they followed the training recommendations, their pupils would manifest increased engagement with the academic material, and ultimately, increased reading achievement. Below we report the findings associated with these hypotheses.

Effects on Student Behavior. High inference ratings on a number of dimensions were made during the posttraining period between January and May. Measures of

attentiveness and passivity are relevant to questions about student on-task behavior. Analyses of variance indicated significant ($p < .10$) teacher training main effects on pupil passivity as measured in the mornings and in the afternoons and on attentiveness in the mornings. Attentiveness in the afternoons showed no significant differences attributable to teacher training. The rank order of the means for passivity was: minimal lowest, then maximal, and then control. Thus, the students in control group classes showed the least involvement in classroom activities, with the same pattern appearing in the morning and the afternoon data. For attentiveness, the rank order was: minimal group highest, then maximal, then control. So these ratings confirm other findings by indicating a detectable treatment effect, with the trained groups' greater implementation apparently resulting in a higher frequency of student involvement in learning activities and more attentiveness to the task at hand.

Effects on Measures of Reading Achievement. Analyses of covariance and regression analyses were performed in an attempt to obtain an unbiased estimate of the size of the treatment effects on achievement measures. Time and space limitations do not permit a discussion of the many complex analyses performed prior to computing the analyses summarized here. For a discussion of issues such as heterogeneity of regression, conceptual bases for the covariate structure, and analyses with the student as the unit, interested readers should see the third volume of the technical report (PTE, 1978).

End-of-year vocabulary scores showed a tendency for a teacher training main effect ($F = 2.078$; $df = 2, 20$; $p = .15$), with the adjusted means being 13.35 for maximal, 13.32 for minimal, and 12.80 for control. The standard deviation of these regression adjusted values was .78, so the difference is about .69 of a standard deviation in favor of the groups receiving training.

There was also a significant ($p < .05$) treatment-by-district interaction that seems due to relatively low scores for the classes in the District I-maximal training group. This same group of classes was at least partly responsible for the lower performance of the maximal training when compared to minimal on the implementation and other measures already discussed. Figure 3 below is the graph of this interaction.

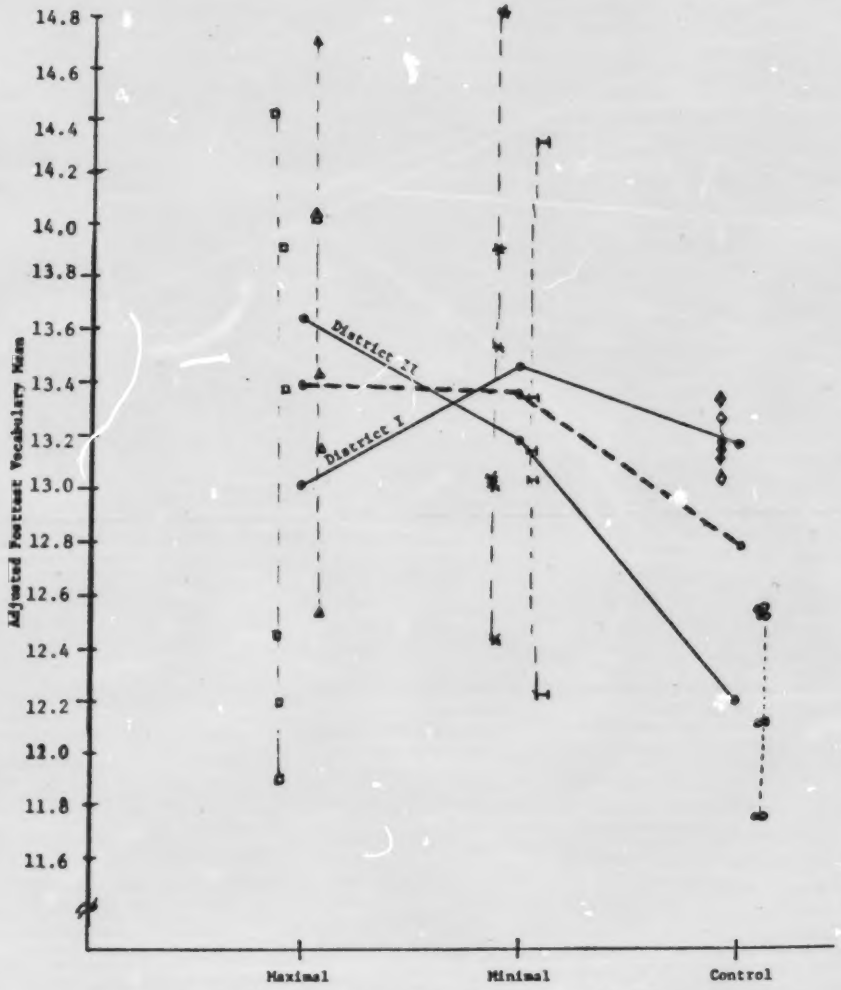


Figure 3. District-by-Teacher Training interaction on adjusted Posttest Vocabulary, N = 33.

Heavy dotted line shows the teacher training main effect means;
□ : District I, maximal classes (N=6)
△ : District II, maximal classes (N=6)
✱ : District I, minimal classes (N=6)
⋈ : District II, minimal classes (N=5)
◇ : District I, control classes (N=6)
⋈ : District II, control classes (N=4)

The results for Reading Comprehension were similar to findings for the Composite Total Reading measure. No significant teacher training main effects were found, but significant interactions emerged (the order of the adjusted means

was: for reading comprehension, maximal, then control, and then minimal; for total reading, maximal, then minimal, then control). Figure 4 below is the plot of the interaction between the teacher training factor and the parent-assisted Learning Skills Program with Total Reading as the dependent variable.

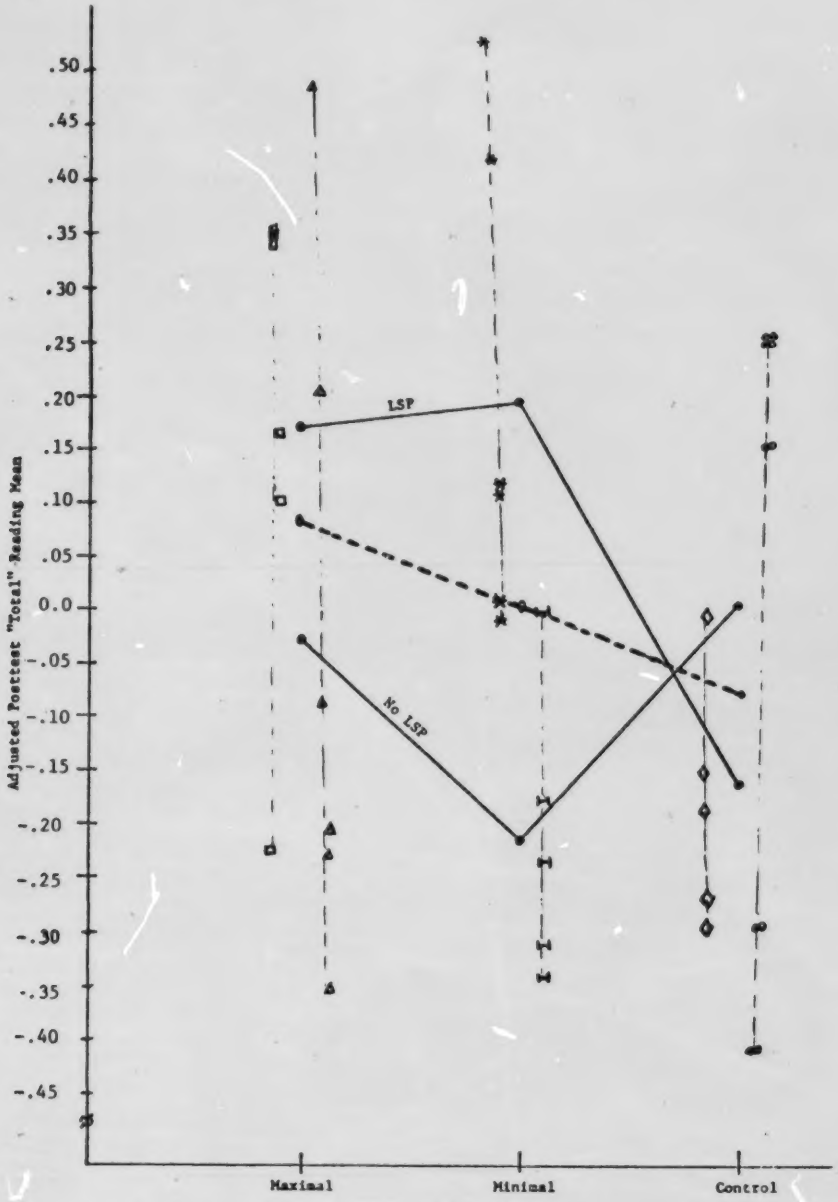


Figure 4. Teacher Training treatment by LSP interaction on adjusted "Total" Reading (z-score composite of Vocabulary and Reading Comprehension), N = 33.

Heavy dotted line shows the teacher training main effect means;
□ : Maximal with the Learning Skills Program (N=6)
Δ : Maximal without the Learning Skills Program (N=6)
✱ : Minimal with the Learning Skills Program (N=6)
⊢ : Minimal without the Learning Skills Program (N=5)
◇ : Control with the Learning Skills Program (N=5)
⊷ : Control without the Learning Skills Program (N=5)

This graph suggests several interesting findings. The main effect means decrease from maximal to minimal to the control group in a nonsignificant linear trend.[1] The groups that received both teacher training and the Learning Skills Program show the highest reading achievement of all the groups (a post-hoc regression comparison revealed those groups to be significantly--$p < .05$--higher than the no-intervention classes).

Unexpectedly, the group of classes receiving neither treatment evidenced relatively high adjusted end-of-year reading scores. The plot of individual class data points suggests that three classes in that group should be studied in more detail--they may have begun the year with higher than average achievement, or those teachers may have been naturally implementing techniques that enhance student achievement.

The design implied by this interaction in Figure 4 is the 6-cell design (3 x 2) that pools classes across districts. Since the design analyzed for heterogeneity of regression slopes was the full 12-cell design, we thought it might be useful to conduct additional analyses of the 6 teacher training by Learning Skills Program groups--in order to follow up on the results suggested by Figure 4. Table 4 below gives the relevant regression statistics.

---

[1] Analysis of covariance of achievement data collected by the school districts (SRA in District I and CTBS in District II) agreed in direction with analyses of the PTE-administered achievement battery, but suggested a more significant difference in favor of the experimental groups on the measure of total achievement. Analyses were conducted separately within districts, and then the significance of the combined results was estimated ($p = .03$). Readers interested in a more detailed discussion of these analyses should see the third volume of the final report (PTE, 1978).

Table 4

Regression Statistics for the 6-Cell Design (pooling across districts) with Total Reading as the Dependent Variable

Full Model: Allows for 6 Intercepts and 6 Slopes

Group	Intercepts	Raw Pre-Post Slopes	95% Confidence Band (± 2 Standard Errors around the slope)
Maximal with LSP	.196	.954	.786 ⟶ 1.122
Maximal without LSP	-.020	.670	.452 ⟶ .888*
Minimal with LSP	.204	1.066	.812 ⟶ 1.320
Minimal without LSP	-.205	.926	.767 ⟶ 1.083
Control with LSP	-.159	.947	.773 ⟶ 1.121
Control without LSP	-.019	1.155	.967 ⟶ 1.343*

Restricted Model: Allows for 6 Intercepts and 1 Pooled Slope

Group	Intercepts
Maximal with LSP	.184
Maximal without LSP	-.026
Minimal with LSP	.198
Minimal without LSP	-.215
Control with LSP	-.172
Control without LSP	-.006

Pooled Pre-Post Slope = .954

F-ratio for the R^2 reduction from the full to the restricted model = 2.464 with 5 and 21 df, and p = .066.

*The 95% confidence bands for these groups' slopes do not span the pooled slope.

These data indicate that the two most extreme slopes (.670 for Maximal training, No Learning Skills Program, and 1.155 for the Control, No Learning Skills Program) are primarily responsible for the near-significant F-ratio suggestive of aptitude-treatment interaction. The high slope for the no-intervention group of classes indicates that their end-of-year reading achievement was heavily contingent on their initial achievement. However, the low slope for half of the maximally-trained teachers (those not receiving the Learning Skills Program) suggests that those teachers were implementing teaching methods during the year that were responsible for their students' end-of-year reading achievement.

Conclusions

We have attempted to present a fairly complete picture of the presently available findings related to the teacher training program. Because of the complexity of the object of the investigation, more time will be required before fully satisfactory analyses and interpretations can be presented. It is possible, at this point, to speculate about the meaning of some of the results.

Soar (1975) discussed the relation between the cognitive level of the outcome and the optimal degree of structure in teaching (as in traditional/structured vs. open/unstructured teaching styles). He stated that

> The sense of the concept is that pupil learning of low cognitive level objectives proceeds best under relatively tightly structured, closely-focused learning conditions, whereas more complex kinds of learning proceed best in settings in which pupils have more freedom to explore and interact with subject-matter, that is, more freedom in their thinking processes (Soar, 1975, p. 14).

Similarly, Stallings (1975) compared the differences in obtained results when achievement is the dependent variable with those obtained when analytic reasoning is the outcome. She observed that

> Higher reading [achievement] scores were obtained in classrooms using systematic instructional patterns where the teacher provides information and asks a question about the information (Stallings, 1975, p. 100),

and that

> This study found that ability to perform well on this [nonverbal problem-solving] test was influenced by the classroom environment and strongly suggests that children who, for a period of three years, have been in classrooms that use a wide variety of activities and provide a wide variety of manipulative materials have learned to see the relationships between parts and wholes (Stallings, 1975, p. 105).

These interpretations also seem to apply to some of the results from this experiment. Our intervention apparently improved vocabulary scores more (or, at least,

with more consistency) than reading scores. If one accepts the idea that vocabulary is an associational construct, as opposed to reading, which might be seen as more integrative and abstract, then these findings make sense in light of the results and interpretations of Soar and Stallings. In many ways the training program recommended somewhat structured teaching--in its emphasis on academic activities and learning, teacher direction and control of events, teacher elicitation of responses, etc.

Some components of the training program were more effectively implemented than others and were more highly related to achievement gains. For example, the Behavior Management and Questioning and Feedback implementation indices were substantially correlated with residual vocabulary and with total reading. The Behavior Management recommendations were concerned with advice to teachers about: using a system of rules to minimize procedural interruptions, avoidance of disciplinary errors (stopping misbehaviors so that they don't spread and increase in severity, targeting disciplinary action at the correct student, and avoiding overreactions), monitoring, and "withitness"--defined as teacher awareness of behavior problems and communication of that awareness to students (see Kounin, 1970). The Questioning and Feedback strategies recommended that teachers use preselection (before asking the question) and calling on non-volunteers as a means of involving all students in public discussions. Other Questioning and Feedback recommendations were that teachers should not encourage student call-outs and should differentiate the difficulty of questions and use of praise and criticism according to the students' characteristic academic orientation.

Together, these variables suggest that "efficient time-use" may be the underlying requisite for achievement-facilitative teaching . Effective teachers proactively avoided behavioral problems and interruptions so that student "time-on-task" could be increased. The questioning methods that they used and that were

positively correlated with achievement were techniques that allow the teacher to maintain control over which students get response opportunities. In discussing their own earlier correlational results, Brophy and Evertson (1974) had maintained that teachers with higher classroom achievement were those who consistently directed questions to all the students in the class (not just letting the children who raised their hands or called out receive the questions).

The results indicating that teacher verbal ability predicted implementation of some training components suggest that the teachers may have served as a "screening" mechanism for the recommendations, with the teachers higher in verbal ability following the recommendations that they believed (by and large, correctly believed) to be effective in enhancing achievement. These findings and other results reflecting the importance of teaching "structuredness" at the beginning of the year imply that future experiments should employ three-dimensional stratification of teachers--one based on class achievement, teacher verbal ability, and the teacher's self-reported structuredness of environment and teaching style--before assigning them to groups (in this study we only stratified on class pretest achievement).

The interaction plotted in Figure 4 is worth additional comment. Higher scores for classes receiving teacher training and the LSP would be expected considering the underlying connection between the two treatments. The LSP was designed for intervention directly with children (via the parents), thus increasing the potential effect of this treatment on measures of student achievement. Further, the purpose of the LSP was to get parents to teach their children how to recognize teacher structuring when it occurs and how to participate effectively in learning-related activities in the class. The data show that, when this kind of treatment is combined with teacher training also intended to promote involvement and impose a degree of structure on the teaching/learning process, substantially enhanced class achievement can be expected.

The data discussed in this paper indicate that, for some outcomes, the additional time and expense of workshop-type training may not be warranted. It is also true, however, as reported by Mitman (1978), that the maximal training teachers in this study felt that a strong point of the training was the opportunity to discuss and practice the strategies with other teachers and with PTE staff, and the minimal training teachers mentioned that they would have appreciated the opportunity for discussion. In short, the attitudes of the maximal training group were more favorable, and teacher attitudes toward training are certainly important considerations in their own right, especially in view of their potential impact on the teachers' long-term level of motivation. It is also possible that, if we had stratified on teacher characteristics before assignment to groups, the differences in results between minimal and maximal training might look quite different. This possiblity is supported by data in Table 3 that show a large portion of the implementation differences between minimal and maximal groups disappearing when adjustments for pretested verbal ability and structuredness are made.

Also, the large amount of observation should perhaps be thought of as part of the teacher training program. The teachers surely knew that one major reason for sending observers to their classrooms was to check on their use of recommended methods. If the minimal group was more inclined to act on this knowledge--in such ways as reviewing their packets the night before the observation--then apparently higher implementation by the minimal group might have resulted.

Additional study of these data should make possible still further improved understanding of what constitutes effective teaching in classrooms.

References

Anderson, L. M., & Brophy, J. E. An experimental investigation of first grade reading group instruction. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976.

Brophy, J. E., & Evertson, C. M. Process-product correlations in the Texas Teacher Effectiveness Study: Final Report. Research Report No. 74-4. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1974.

Brophy, J., Evertson, C., Crawford, J., & Sherman, G. The Texas Teacher Effectiveness Study: Student sex, grade, and socioeconomic status differences in classroom process measures. Report No. 75-21. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1975.

Corno, L. Effects of parent instruction on teacher structuring and student participation in the third grade: An aptitude-treatment interaction approach. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

Kounin, J. Discipline and group management in classrooms. New York: Holt, Rinehart & Winston, Inc., 1970.

McDonald, F., & Elias, P. The effects of teaching performance on pupil learning, Beginning Teacher Evaluation Study: Phase II, Final Report: Vol. I. Princeton, N. J.: Educational Testing Service, 1976.

Mitman, A. L. Teacher self-report in an experiment on teacher education and effectiveness. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

Program on Teaching Effectiveness, CERAS. Final report on the experiment on teacher effectiveness and parent-assisted instruction in the third grade: Volume 1, February 1977.

Program on Teaching Effectiveness, CERAS. Final report on the experiment on teacher effectiveness and parent-assisted instruction in the third grade: Volume 3, 1978.

Ryans, D. Teacher characteristics schedule, revised form G 70/2. Teacher Characteristics Study II-2. Honolulu, Hawaii: U.H. EDRAD, 1970.

Soar, R. Final report, Follow Through classroom process measurement and pupil growth (1970-1971). Gainesville, Fla.: Institute for the Development of Human Resources, University of Florida, 1973.

Soar, R. An integration of findings from four studies of teacher effectiveness. Paper presented at an invitational conference on research on effective teaching sponsored by NIE and the California Commission on Teacher Preparation and Licensing, San Diego, August 1975.

Stallings, J. Implementation and child effects of teaching practices in Follow Through Classrooms. Monographs of the Society for Research in Child Development, 1975, 40 (7-8, Serial No. 163).

Stallings, J., & Kaskowitz, D. Follow Through classroom observation evaluation, 1972-1973. Menlo Park, Cal.: Stanford Research Institute, 1974. Office of Education Contract OEC-0-8522480-4633(100).

Stayrook, N., and Crawford, J. An experiment on teacher effectiveness and parent-assisted instruction in the third grade: The observation data. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

Appendix A:

Summary (Fifth) Teacher

Training Packet

Teacher Training Packet 5: A Summary and Review

Program on Teaching Effectiveness

Stanford Center for Research and Development in Teaching

Fall 1976

Over the past five weeks you have been receiving materials describing the ways of teaching that four recent studies have indicated are associated with higher pupil achievement. The purpose of this packet is to review and summarize those teaching strategies. Since we restricted the content of our in-service training to methods and activities that research has shown to relate to higher pupil achievement, we have no doubt excluded many other aspects of "effective teaching." The reasons for this are twofold:

(a) We did not want to introduce methods not based on the previous observational research, since it would not then be possible to attribute any improvement in pupil achievement (expected in the classrooms of the trained teachers) to the methods suggested by the research studies.

(b) Other criteria (in addition to higher pupil achievement) were not used because of a lack of a firm basis in the earlier research findings. There are many opinions about what constitutes good teaching for the promotion of affective outcomes (self-esteem, higher motivation, lower anxiety, etc.). But there is a scarcity of actual evidence about the teaching dimensions that are associated with higher pupil self-esteem, motivation, or freedom from anxiety.

It might have been better to have devised a more "complete" or inclusive training course, yet we feel that those lost benefits must be weighed against what we will gain from establishing, for the first time

in the field of research on teaching, that the training of teachers on methods associated with pupil achievement actually causes improved academic learning. We also feel justified in focusing on academic outcomes because increases in proficiency in this area may be expected to be related to better pupil attitudes toward themselves and academic activities. Also, our society has recently become even more concerned about the academic effectiveness of public schools. (Scholastic aptitude scores of high school students entering college have steadily declined over the past ten years.) The basic academic skills are typically acquired in the first three or four years of schooling. Our study is being conducted at the third-grade level to see whether in-service teacher training on selected teaching strategies can have a positive effect on pupil academic growth at an important time in the young child's educational career.

We will now first summarize briefly the strategies presented in the series of training packets; then on the following pages these strategies will be further discussed and elaborated.

* * * * *

Behavior Management and Classroom Discipline

(1) Teachers should have a system of rules that allows pupils to attend to their personal and procedural needs without having to check with the teacher.

(2) Teachers should prevent misbehaviors from continuing long enough to increase in severity or spread to and affect other children.

(3) Teachers should attempt to direct disciplinary action accurately--that is, at the child who is the primary cause of a disruption.

(4) Teachers should keep "overreactions" to a minimum (even though over-reactions are probably effective in stopping the misbehavior).

(5) Teachers (and aides, if present) should move around the room a lot, monitor pupils' seatwork, and communicate to the pupils an awareness of their behavior, while also attending to their academic needs.

Instructional Methods

(6) When pupils work independently, teachers should insure that the assignments are interesting and worthwhile and still easy enough to be completed by each third-grader working without teacher direction.

(7) Teachers should keep to a minimum such activities as giving directions and organizing the class for instruction. They can do this by writing the daily schedule on the board, insuring that pupils know where to go and what to do, etc.

(8) Teachers should spend at least one-third to one-half of their time teaching larger groups of pupils (more than eight children). When they do teach smaller groups or individuals, they should take steps to make sure that the other pupils in the class have work to which they can attend.

(9) Teachers should make abundant use of textbooks, workbooks, and other pencil-and-paper activities. These have been found to be associated with higher pupil achievement. But the use of games, toys, and machines has not been found to be associated with higher pupil achievement.

(10) Teachers should provide visual demonstrations and phonics exercises in conjunction with reading activities.

(11) Teachers should frequently conduct public (i.e., addressed to a larger group or the whole class) question-and-answer sessions concerned with the academic subject matter at hand. With less academically oriented pupils, teachers may find it helpful to initiate some brief private discussions concerning personal matters.

Specific Methods for Asking Questions and Providing Feedback

(12) In selecting pupils to respond to questions, teachers should use the technique of calling on a child by name before asking the question, as a means of insuring that all pupils are given an equal number of opportunities to answer questions.

(13) Teachers should avoid calling on volunteers more than 10 or 15 percent of the time during question-and-answer sessions. It is also advisable to discourage pupil "call outs" to questions asked of other children (except possibly from less academically oriented children who may benefit from this type of activity).

(14) In the interest of promoting smooth, task-oriented discussions, teachers should not encourage large numbers of pupil-initiated questions and comments. It is also important for teachers to listen carefully to pupils' opinions and, if a disagreement is called for, to express such disagreement to the child.

(15) With less academically oriented pupils, teachers should ask easier questions--questions that can almost always be answered correctly. When questioning more academically oriented pupils, teachers should ask more difficult questions--questions that are answered incorrectly about one fourth of the time.

(16) Teachers should give praise only for really outstanding work; also, praise is likely to be more effective with less academically oriented pupils. Mild criticism is effective in communicating higher expectations ("you can do better") to more academically oriented pupils.

(17) With less academically oriented pupils, teachers should always aim at getting the child to give some kind of response to a question. Rephrasing, giving clues, or asking a new question can be useful techniques for bringing

forth some answer from a previously silent pupil or one who says "I don't know" or answers incorrectly.

(18) With more academically oriented pupils who generally become actively involved in discussions, teachers should concentrate on getting the correct response. Therefore they should redirect questions to other pupils if the more academically oriented pupil answers incorrectly.

(19) Teachers should give the answer (to both more and less academically oriented pupils) if the response is at least partly correct. Teachers should not simply repeat the same question if any pupil (either more or less academically oriented) answers incorrectly, says "I don't know," or remains silent.

(20) With more academically oriented pupils, teachers should give brief feedback extensively (80% or more of the time) during private, one-to-one discussions. When dealing with less academically oriented pupils, teachers should use approximately equal amounts of brief and longer feedback, tailoring the duration of their reactions to the needs of the individual child in each situation.

(21) During reading-group instruction, teachers should give a maximal amount of brief feedback, and provide fast-paced activities of the "drill" type.

(22) During public question-and-answer sessions, teachers should occasionally give a detailed, "why" explanation in answer to a question.

* * * * *

Together these recommendations imply that a classroom should run efficiently and smoothly, display a high degree of "task-oriented" behavior, have few serious or recurrent behavior problems, and generally may be expected to show higher achievement gains. While these methods and activities

do not include all the possible aspects of effective teaching, they are the ones most firmly grounded in research findings. We have strong reasons to believe that, if you can use all or most of these recommended methods in your classroom in a manner compatible with your preferred way of teaching, your pupils should make greater achievement gains than would otherwise occur.

The methods and behaviors outlined under the Behavior Management and Classroom Discipline heading indicate that highly effective classroom managers were active in anticipating and avoiding behavior problems, rather than being content to wait for a misbehavior to occur and then reacting (and hence providing the opportunity for a disciplinary error to occur). Having a well-known system of rules, frequently monitoring pupils' seatwork, and selecting materials which the pupils can and will attend to--all these are valuable ways of reducing the frequency of discipline problems.

The concept of "withitness" mentioned in the second set of training materials is worth returning to in this summary discussion. "Withit" teachers are highly aware of the activities going on in their classrooms and also effectively and explicitly communicate this awareness to their pupils. Withitness is facilitated by moving around the classroom and monitoring seatwork. Teachers' actions should be aimed accurately at the pupil who is the main cause of a disturbance and should stop the misbehavior before it continues too long. These teacher actions will let the pupils know that the teacher "really knows everything that is going on." An increase in a teacher's "withitness" is associated with fewer behavior problems in the class, and accordingly with more time available for the direct academic instruction that apparently promotes learning at the third-grade level.

The ways of teaching discussed in the Instructional Methods section all

have as an underlying theme the notion that teachers who maximize academic, "on-task" activities have third-grade classes that show greater gains in achievement. This theme also applies to the recommendations about avoiding spending too much time in direction-giving and organizing, and teaching to larger groups of pupils. The most effective teachers are those who make efficient use of available time in meeting the educational needs of all of their pupils.

We recommend frequent publicly-held question-and-answer sessions as a vehicle for getting the children involved in learning-related activities. This activity is also useful for giving pupils the opportunity of learning and benefiting from what their fellow pupils have to say. We also make fairly specific recommendations in favor of using texts, pencil-and-paper activities, visual demonstrations, and phonics exercises since these ways of teaching were found to occur more often in classrooms that tended to have higher achievement scores.

While the recommendations concerning the types of instructional methods found related to higher classroom achievement <u>are</u> obviously academically oriented, this orientation should not be taken to mean that we are unconcerned with affective outcomes. As we have argued elsewhere in these training materials, by increasing the pupils' mastery of academic objectives, we may expect them to have better attitudes toward school, feel less anxious about academic matters, and increase their perceived self-worth.

The fourth set of training materials contains extensive information about fairly specific ways of <u>Questioning Pupils and Providing Feedback</u> following their responses. One principle--underlying the recommendations about calling on pupils before asking questions, avoiding calling on

volunteers, and discouraging "call outs" (unless a less academically oriented pupil is calling out an answer)--is that the more effective teachers maintain control over which pupil responds. Furthermore, such teachers use this control in making sure that the opportunities to answer questions are equally spread over all pupils--not just those eager learners who are always willing to volunteer and answer questions.

Most of the other recommendations concerning questioning and feedback strategies are differentiated according to the type of pupil being taught:

For Less Academically Oriented Pupils	For More Academically Oriented Pupils
• Ask easier questions which can almost always be answered correctly.	• Ask more difficult questions which are answered incorrectly about one-fourth of the time.
• Give praise for outstanding work; avoid the use of criticism.	• Give praise very sparingly and appropriately; use criticism for communicating higher expectations.
• Do not allow the child to remain silent when asked a question; use rephrasing and give clues to get some response.	• Concentrate on getting the correct response; redirect question to another pupil if the first pupil's answer is incorrect.
• In facilitating pupil responding, do not actively discourage "call outs."	• Tell pupils not to call out answers, but to wait for their turn.
• Use equal amounts of brief and longer feedback, according to the demands of the situation and the needs of the pupil.	• Use brief feedback extensively (80% or more of the time), but interact for longer periods when this is needed.

Not all of the recommendations on questioning and feedback are different for the two types of pupil:

For Both More and Less Academically Oriented Pupils

- Give the correct answer if the child's response is partly correct.
- Avoid simply repeating the question if the child answers incorrectly, says "I don't know," or makes no response.
- Keep reading-group activities highly task-oriented and fast-paced; use brief feedback for this purpose.
- During public questioning sessions, occasionally give detailed "why" explanations of answers.

We believe that the recommendations about treating the two types of children differently (to bring them both to the same goal of greater academic proficiency) are defensible in the light of the differences in their behavioral characteristics, as described in the first training packet. In a sense, it's a matter of being aware of what the child would normally "get" or "expect" in the way of instruction and supplementing that experience with the needed "extra" help, direction, or interaction. The less academically oriented child is not too interested in academic matters, has probably not been praised very much, and may not have had much success in answering questions in class. The idea, therefore, is to get this pupil involved in the learning activities, praise him or her when such involvement is shown, and ask questions that are not too difficult. The more academically oriented pupil is typically already interested in the subject matter; so the teacher's task is to provide more challenge and increase his or her motivation (by asking difficult questions and making some "critical" comments) while still preventing this child from dominating the classroom learning activities. While it would theoretically be desirable to construct individual profiles for each child in your class and then interact with each child in a uniquely appropriate way, we hope that our ideas about the best ways of teaching these two general "types" will serve as a practical aid in individualizing

your instruction.

Time and next year's evaluation of our program will tell whether this effort succeeds in improving pupil achievement. Any credit for such a positive outcome must first go to you, as a teacher who unselfishly gave time out of a schedule that had little time to spare. Thank you once again.

Effects of Parent Instruction on Teacher Structuring and Student Participation in the Third Grade: An Aptitude-Treatment Interaction Approach

Lyn Corno

Stanford University

School of Education

Center for Educational Research at Stanford

Program on Teaching Effectiveness

Abstract

The study assessed the effects of a parent-assisted Learning Skills Program (LSP) in the context of an experiment on teacher effectiveness. Based on previous research, the LSP was designed to develop cognitive and behavioral processes that "successful" students use in adapting to school. The curriculum model followed a "technological" approach, focusing on concept acquisition and practice on appropriate reading material.

For six weeks in the Fall of 1976, LSP program exercises were distributed to 416 third grade students in 33 classes that were stratified on pretest ability and randomly assigned to treatments. Effects were assessed at two stages with a number of aptitude-outcome and process-level measures. Stage I determined the incidence of naturally occurring aptitude-treatment interactions (ATI) on which the LSP materials were based. This was done to ascertain the validity of the LSP treatment in the present sample. Stage II assessed LSP treatment and ATI effects.

Stage I results produced naturally occurring ATI at the class level of analysis that were psychologically meaningful and consistent with previous research. In Stage II it was found that LSP group students scored significantly higher than controls on an achievement test covering LSP concepts--a result that was maintained on the same test readministered four months later. Additionally, reading achievement was significantly higher for LSP students than for non-LSP students, particularly when program completion rates were high. Interaction effects were interpretable in terms of different program completion rates across treatment groups. Attrition was related to student general ability; to teacher characteristics, including program support; and to parent instructional variations. Three sizable LSP-related ATI were found at the class level. These appeared reasonably stable despite the small number of classes and were also interpretable and consistent with prior research.

Effects of Parent Instruction on Teacher Structuring and Student Participation in the Third Grade: An Aptitude-Treatment Interaction Approach

The Research Problem and Design

Traditionally, the task of the educator has been to see that learning, pleasure, and rewards predominate in school. That teachers often fail to perform this task is a point few would dispute. Still, there are students who manage to learn with pleasure and reward quite on their own. For one reason or another, some students are more adept than others at determining what it takes to do well in school. These more adept students may be said to have acquired a set of "learning skills" that somehow work.

Recent accumulating research evidence supports the need to attend to student individual differences in instructional situations. In turn, a theory and method for systematic inquiry into the process of accommodating individual differences has been formalized as aptitude-treatment interaction (ATI) research (Cronbach & Snow, 1977). The basic hypothesis of ATI research is that students high in some characteristic (aptitude) will respond better under one instructional method (treatment), and students low in that characteristic will respond better under another. Aptitude is defined as any student characteristic that forecasts instructional outcomes.

This study was conceived and conducted within the framework of ATI research. The main purpose of the study was to determine whether a set of classroom learning skills for third graders identified on the basis of ATI research findings would help students learn. These learning skills, formulated specifically in a systematic instructional program (hereafter called the Learning Skills Program, LSP), represented an attempt to make explicit the kinds of cognition and behavior that "successful" students use to adapt to the world of the classroom. Accordingly, the major goal of the LSP was to provide students with a set of procedures for coping with third-grade performance

requirements, in hopes of easing the burden of information processing and classroom participation demands on students. Table 1 lists the LSP objectives and the component program exercises, grouped into two general categories: understanding teacher structuring (or organizing information) and participation. These two kinds of student comprehension and skills

---

Insert Table 1 about here

---

have been shown to relate to increased academic achievement, depending on the verbal aptititude and anxiety level of individual students (Snow, 1976).

The study, conducted in two phases, was part of an experiment on teacher training and effectiveness (Program on Teaching Effectiveness, PTE, 1978). Phase I was an exploratory ATI analysis intended to yield hypotheses about effects of the LSP. Actual effects of the LSP were with a number of measures of cognitive and motivational variables (shown in Table 2), and various kinds of data on classroom processes. All LSP materials

---

Insert Table 2 about here

---

were administered to students by their parents in the home. This procedure permitted assessment of LSP effects as distinct from teacher variables and also provided a novel form of parent involvement.

Hypotheses

The hypotheses were based on three bodies of research. Phase I hypotheses were based on ATI results concerning selected cognitive and motivational aptitude variables and structuring and participating treatment variables (after Snow, 1976). Phase II hypotheses were based in part on the results of Phase I

and in part on the findings of (a) several large-scale teaching effectiveness studies and (b) previous research on parent-assisted instruction.

Expected Results, Part I

1. Greater amounts of teacher structuring should be especially beneficial for high-anxiety students, and possibly also for students with low/low and high/high ability-anxiety combinations.

2. As teacher demands for participation increase, achievement should increase for low-anxiety students. The high-anxiety students, in contrast, should do best when participation demands are low. In a three-dimensional graph, taking ability into account, one would expect a picture of a plane with a high point in the upper right back corner (high ability, low anxiety). Thus ability would mediate the effects of anxiety on achievement under high participation conditions.

Results such as these would suggest the first hypothesis regarding the effects of the LSP listed below. Hypotheses 2 and 3 of Part II are based on the Teaching Effectiveness and Parent Involvement research.

Expected Results, Part II

1. The LSP should be especially beneficial for students of low ability and high anxiety or those of high ability and low anxiety in classes where teacher structuring and teacher demands for participation are low.

2. Students who complete the LSP should manifest greater understanding and awareness of teacher structuring and participating behavior than students not completing the LSP. It is also possible--although not directly hypothesized--that these students will increase their classroom participation as a result of LSP influence.

3. Students who complete the LSP should obtain higher adjusted reading achievement scores than students in the comparison group.

Even partial support for these hypotheses should provide important knowledge about the usefulness of attempting to help third graders adjust to schooling practices.

Method

The LSP was represented as a two-level factor in the 2 X 3 experimental design of the larger study. This design was replicated in two school districts, so that 17 out of 33 sample classes received the LSP, contrasted with 16 that did not. The three-level teacher training factor distributed the "LSP" and "no LSP" classes over three teacher training groups: "maximal," "minimal," and "control," respectively (see PTE, 1978, for more detailed information). A total of 840 third-grade students participated in the study.

Program Development

For presentation to students, the organizing skills were defined as "tricks for making ideas orderly." Similarly, the participating skills were defined as "tricks for sharing your ideas" (refer to Table 1). The exercise format followed a concept acquisition strategy, where the objective was that the student be able to identify instances of the concept as it appeared in mock classroom discussions. Each exercise began with a brief review and a statement of the goal; was accompanied by a "trick card" listing the important points; and concluded with a summary. Thus the methods which the students were being taught to recognize and use were actually incorporated into the exercise sequence.

The LSP was based on a "technological" approach (Eisner & Vallance, 1974), involving prespecified objectives, appropriate practice, and knowledge of results. The eight skills were presented in the form of classroom scenarios, and in each case, the exercise contained a certain amount of reading practice material as well. Pilot tests were conducted and comparisons were made with other third grade instructional material to insure the appropriateness of all vocabulary words except the key concepts to be taught. Each exercise was designed to require about 30 minutes. A sample LSP exercise is contained in Appendix A.

Procedure

The independent "treatment" variables used in the ATI analyses for Phase I consisted of time-sampling measures of teacher structuring behavior and teacher demands for participation. Measures of these variables were obtained for each teacher with a modified version of a classroom observation instrument that permitted coding the relevant behavior in numerous five-minute time-samples during 16 full school days of observation (Stallings & Kaskowitz, 1974). Four five-minute observation segments were also collected on each student in the study. These "low inference" classroom interaction data were supplemented with several "high inference" ratings made by the observers at the end of each half-day of observation.

For six consecutive weeks in the Fall of 1976, students in the LSP group took the exercises home to complete with their parents. Participants were asked to complete two 30-minute exercises each week on a voluntary basis. This procedure required parent-child teams to spend approximately one hour a week on LSP material at home, so that students completing the full set of LSP exercises received approximately five hours of instruction (containing reading practice material as well) over and above that given in school. Warm-up and follow-up exercises were also provided to acquaint participants with the program's format and procedure and to assess student and parent attitudes toward the program.

To standardize delivery, the exercises were scripted, and parents were encouraged to follow the scripts. With each exercise, a checklist was also provided for the parents to use in monitoring the child's progress, providing feedback to the investigator, and reporting the degree to which they had felt it necessary to deviate from the script. With the fifth exercise a letter was distributed, which provided responses to major parent checklist comments.

Assessment was done with three criterion-referenced tests, occurring at strategic points in the program, and a final, 15-item "yes-no" achievement test. The last-named instrument was used to measure knowledge of program concepts in all students in the sample. The criterion-referenced tests were distributed only to students who completed the program, and were intended for use as supplementary learning devices rather than valid measures of achievement. The 15-item achievement test was administered by the research staff as part of the batteries of mid-year and end-of-year tests. Also, before the LSP began, all students completed an aptitude test battery administered by the research staff (see Table 2). The mid-year and end-of-year test batteries contained parallel forms of the reading achievement test given at pretest.

Major Results and Discussion

Implementation

Several factors affected program implementation, resulting in a considerable attrition problem. Of the 416 students originally assigned to the LSP group, only 136 (32%) completed the entire eight-exercise sequence. The mean number of exercises completed across all 17 LSP classes was 6.04, with a standard deviation of 2.39. Since the LSP was presented on a purely voluntary basis, parents were free to discontinue the exercises any time. Such voluntary participation was considered necessary to the maintenance of good working relationships with the schools and parents.

Path analyses of the 17 LSP class means on LSP Achievement and Reading revealed several factors that influenced decisions to discontinue the program. Among these were student general ability, teacher support, teacher concern about rapport with students, and the extent to which parents deviated from the intended script. Figure 1 presents results of the path analyses using LSP Achievement test means as the dependent variable.

Insert Figure 1 about here

Where reasons were given for decisions to discontinue, parents frequently indicated that the exercises were too difficult and abstract for the children. Indeed, LSP completion rates were significantly and positively related ($p < .05$) to the composite student aptitude measure defined as "G" (or general mental ability--a combination of standardized pretest scores on (a) reading and vocabulary, which represent crystallized ability, or G_c, and (b) the Raven's Matrices test, which represents fluid ability, or G_f, Snow, 1976). The correlation was .42, with a total class N of 17; but, as Figure 1 indicates, the direct causal influence of G was mediated by teacher support. Thus it appeared that the LSP was more likely to be completed by higher ability students unless the teacher was quite encouraging.

LSP teachers varied considerably in degree of support for the program. While no effort was made to provide recommendations to teachers for program support, teacher interview data indicated that some teachers made an effort to praise the program, dispense explicit rewards, and provide other motivators, such as substituting the exercises for reading homework. Among these "highly supportive" teachers (as rated on a 3-point scale from low to high), completion rates (defined as percent of students completing the entire program) ranged from 38% to 58%. Teachers providing "neutral support," with no explicit motivators, had completion rates ranging from 16% to 30%. The correlation between teacher support and mean class LSP completion percentage was .66 ($N = 17$).

Of the 136 families completing the entire program, 68% found the LSP "quite to moderately difficult" for their child. More than a third (43%) of the families felt the program had increased their child's class participation. And finally, 61% and 62% of these parents indicated the parent and the child, respectively, felt positive or very positive about the program.

Effects on Students--Treatment Effects

Analyses of treatment effects were conducted using a complex covariance adjustment procedure adapted by PTE from suggestions in Cronbach, Deken, and Webb (1976). When the assumptions of the model are satisfied (in particular, homogeneity of variance and regression), the procedure can be used at the individual level of analysis to perform a "double" adjustment for aptitude (in this case, the aptitude measure was the composite variable labeled G). The adjustment is double in that it removes from each individual student's score the cohort or "group" effect due to class membership, as well as the individual's aptitude deviation score (individuals' scores deviated from their class mean). In the class level analysis (using an N of 33) only the "between class" or class membership adjustment is made.

Analyses of treatment effects were performed only on the six cognitive outcome measures (the January and May LSP Achievement tests, January and May Reading tests, May Vocabulary test, and May Raven's Matrices test). G was the covariate in each case and the double adjustment was made where the above-mentioned assumptions were satisfied in the data. Analyses of covariance were performed at the class and the individual level, in the first instance with the full experimental design (12 cells--2 Districts x 3 Training groups x 2 LSP groups), and in the second with a post hoc design defined by the two-group district factor and a four-group factor (labeled "levels"), which represented the degree of LSP completion. Level 1 students completed 0 or 1 exercises; level 2 students, 2 to 6 exercises; and level 3, 7 or 8 exercises. All non-LSP control students were labeled level 0. Thus the post hoc design was an 8-cell configuration for which the same covariance adjustments were made (where the assumptions were met). Analyses with this design were performed to provide a more precise assessment of LSP effects.

Major analysis results are summarized in Table 3. Interpretations follow:

Insert Table 3 about here

---

1. Despite the attrition problem, LSP-group students and classes scored significantly higher than controls on the LSP Achievement test given immediately after LSP completion, on the same test given four months later, and on the vocabulary test. Where interaction effects were found, results were interpretable in terms of variation in completion rates across the three teacher training groups and two school districts. This variation was at least partially explainable by teacher support and student ability differences.
2. With the LSP levels distinction, significant or near-significant LSP effects appeared on both LSP Achievement measures and the May reading measure. These effects appeared at the student level only, in each case, except for the immediate LSP test, on which the effect was in evidence at the class level also. Again, the district by levels interaction was influenced by the different completion rates in the two districts. These rates covaried with ability differences of students and teacher support.
3. The LSP group was superior to the control group on a standardized composite outcome termed "total reading" (which combined May reading and vocabulary scores). Treatment by LSP and district by treatment by LSP interaction effects were interpretable in terms of treatment and district rather than LSP group differences (see PTE, 1978). No sizable LSP treatment or interaction effects were found on the remaining outcome measures of January reading or Raven's Matrices.

Effects on Students--ATI Effects

ATI were investigated using generalized regression analysis procedures recommended by Cronbach and Snow (1977). Aptitudes were, in each case, standardized deviation scores. (Each student's score was converted to a deviation from his class mean and class means were converted to a deviation from the grand mean.) The aptitudes were defined as G, G_d (an orthogonal contrast representing the difference between G_c and G_f), and A_x (total student anxiety). Regression analyses were conducted at the individual level (using a pooled within-class regression line) and the class level (as recommended by Cronbach, Deken, and Webb, 1976) for each of the previously specified six cognitive outcomes. Aptitudes and aptitude-by-aptitude combinations (multiplicative aptitude terms), treatment main effects, and ATI were entered into the equations in a predetermined order, according to theoretical expectations and the results of previous research. Initial analyses were considered exploratory and were used to specify refined models for more powerful secondary tests.

Results of the ATI analyses using observation data on teacher structuring and participating as continuous treatments produced a number of ATI at both the class and student levels. These effects were interpretable and conformed to results of previous research concerning the complex relationship among the structuring-participation treatments and the ability-anxiety aptitudes (except that the previous research results were found at the individual level of analysis). For example, classes with opposite pattern aptitude combinations (High Ability and Low Anxiety, and vice versa) did best on May reading achievement when the teacher tended to be less structured and less demanding of participation. This finding was altogether similar to Peterson's (1977) result for the achievement test scores of individual high school students.

Since ATI effects did appear in the present sample, still more complex ATI effects were expected in analyses using the independent variables (LSP, Training, and District) along with observers' codings of teacher structuring and participation demands as treatments. This was the case, and again ATI were found at the class and student level. The fact that so many class level ATI did result was especially interesting because ATI studies have typically examined effects at only the pooled within-class level.

Cronbach and Snow (1977) recommended sample sizes of 100 or more per treatment in analyses of between-class regressions since the sampling error of a regression slope based on less than 100 class means is usually large. So, in order to further examine the nature of the ATI, 95 percent confidence intervals were computed for the regression lines in the present sample of 33 classes. For six of the eight sizable ATI effects, the 95 percent confidence band did not span zero. However, for only three of these did the ATI account for a percentage of variance that was significant or near-significant. These three effects alone were taken seriously in resulting interpretations and conclusions.

Table 3 also summarizes the eight primary ATI effects. Interpretations are as follows:

1. Figure 2 illustrates the significant ATI (G x A_x x LSP) on May LSP Achievement. LSP-group classes of low average ability and anxiety

---

Insert Figure 2 about here

---

did particularly well on the LSP test. Non-LSP classes of students with high ability and high anxiety were able to perform well on the test despite no instruction in the LSP concepts. The LSP was also

slightly more beneficial for high ability-low anxiety classes; but for low ability-high anxiety classes the program made no difference. Thus it appeared the LSP was most useful, in terms of improving knowledge of structuring and participating concepts, for classes with low ability and low anxiety. Perhaps the LSP taught these classes something they did not already know, making their performance almost equivalent to that of the brighter, more anxious classes who had somehow grasped these concepts without formal instruction.

2. Also, the LSP appeared to benefit high anxiety class groups when vocabulary was the outcome. Though the LSP-group slope for vocabulary on anxiety was nearly flat, the No-LSP group slope was moderately negative, suggesting that high anxiety classes may have suffered somewhat from a lack of program support.
3. Finally, the interaction between ability, anxiety, and LSP in influencing nonverbal reasoning as a dependent variable produced a picture in which all aptitude-pattern groups except high ability/low anxiety classes did best <u>without</u> the LSP. For classes of high ability-low anxiety students, either treatment produced about the same moderately high result. One possible explanation for this finding is that the LSP was a highly verbal program that required, above all, concept acquisition, reading, and an understanding of the purpose of teacher verbalizations. It may be that, because of this strong emphasis on verbal fluency, analytic reasoning suffered for all students but those who were strong enough in that area to begin with and were also unhampered by high anxiety. A

second possible explanation may be that the LSP classes, by participating in this structured, verbal treatment, were somehow engaging in less of the sort of activity required to develop analytic reasoning, such as object manipulation, conceptualization, and so on. In short, it is possible that the parental attention and teacher encouragement received by students during the LSP treatment could have helped to engender in these class groups a cognitive "set" for verbal learning. These explanations cannot be evaluated from the data at hand. And, once again, all of the ATI effects are contaminated by variations in completion rates across classes. Still, such results and speculation can lead to interesting hypotheses for future research.

Conclusions and Recommendations

The data indicate that the LSP treatment had a complex pattern of effects in these third grade classes. In brief, the program improved reading comprehension and vocabulary, particularly when all of the exercises were completed. To the extent that the number of exercises completed by a student can be considered a measure of content covered, the added effort expended in the verbal learning area probably made the difference. This difference was hypothesized (as noted under Part II on page 3), but an absolute test of the hypothesis would require a "clean" application of the program (to eliminate the attrition).

In addition, the LSP increased student understanding of teacher structuring and class participation, despite the attrition problem. This understanding was substantially retained for four months following instruction. Student, teacher, and parent variables mediated these outcome effects. While the path

analyses based on 17 class means must be interpreted with caution, further research designed to explore the influence of these or similar factors should prove helpful in reducing attrition problems in future implementation efforts. The fact that the LSP was generally well regarded by the 136 families, who completed all eight exercises, suggests that the idea is already fairly appealing and can be made more so by further developmental work.

The LSP did have complex effects, depending on class ability and anxiety configurations. The program did appear beneficial for classes of students with "opposite-pattern" levels of verbal ability and anxiety when teacher structuring and participation demands were low (see the hypothesis, page 4), as evidenced by an inspection of class mean scatterplots for the appropriate group. The major between-class ATI effects, though interesting and psychologically meaningful, will need to be reassessed in a second sample of classes to ascertain their validity.

And finally, the primary purpose of the study--to provide instruction that might help students of minimal aptitude adapt to the world of the classroom--was fairly well achieved. Low-ability, low-anxiety classes, in particular, seemed to benefit from instruction in LSP concepts (see Figure 2). There was also indication that high-anxiety classes miss an opportunity for improved vocabulary performance without the LSP experience.

Perhaps more investigations of instructional methods would find it profitable to adopt this strategy of improving pupils' learning skills. It is certainly more optimistic than the alternative of accepting aptitude deficiencies as immutable and adapting the teaching method instead. While both approaches to instructional adaptation have a place, the former has been far less common in research and in practice. For that reason alone it deserves further exploration.

A second test of the LSP is planned for the future, following program revisions. The study will be conducted at a higher grade level in an attempt to minimize attrition. In addition, teacher and parent training will be provided to increase support. It is hoped that this second tryout, by eliminating some of the problems of the present study, will provide additional useful information on LSP effects.

References

Cronbach, L. J., Deken, J. E., & Webb, N. Research on classrooms and schools: Formulations of questions, design, and analysis. The Stanford Evaluation Consortium, Occasional Paper. Stanford, Ca., July 1976.

Cronbach, L. J., & Snow, R. E. Aptitudes and instructional methods: A handbook for research on interactions. New York: Irvington/Naiburg, 1977.

Durost, W. N., Bixler, H. H., Wrightstone, J. W., Prescott, G. A., & Balow, I. H. Metropolitan Achievement Test, Form F. Test 2: Reading. New York: Harcourt Brace Jovanovich, 1970. Pp. 4-7.

Eisner, E. W., & Vallance, E. (Eds.) Conflicting conceptions of curriculum. Berkeley, Ca.: McCutchen, 1974.

Peterson, P. L. Interactive effects of student anxiety, achievement orientation, and teacher behavior on student achievement and attitude. Journal of Educational Psychology, 1977, 6, 779-793.

Phillips, B. N. Children's school questionnaire (short form). Austin: University of Texas, 1966.

Program on Teaching Effectiveness. An experiment on teacher effectiveness and parent-assisted instruction in the third grade: Vols. I-III. Stanford, Ca.: Center for Educational Research at Stanford, 1978. (National Institute of Education Contract No. NE-C-00-0061)

Ravèn, J. C. Guide to the standard progressive matrices, Sets A, B, C, D, and E. New York: The Psychological Corporation, 1960.

Snow, R. E. Research on aptitude for learning: a progress report. Stanford, Ca.: Stanford University School of Education, 1976. (Spencer Foundation Grant and Office of Naval Research Contract No. N0014-75-C-0882)

Stallings, J. A., & Kaskowitz, D. H. Follow Through classroom observation evaluation 1972-1973. Menlo Park, Ca.: Stanford Research Institute, 1974, U. S. Office of Education Contract OEC-0-8522480-4633(100).

Thorndike, R. L., & Hagen, E. Examiner's manual. Cognitive abilities test. New York: Houghton Mifflin, 1971.

Table 1

Discrete Skills Constituting the Composite Independent Variables of Structuring and Participating

Composite Independent Variable	Discrete Skills	
	Skill Name	LSP Exercise Number
Structuring ("Making Ideas Orderly")	Review	1
	Goals	2
	Important Points	3
	Summary	4
Participating ("Sharing Your Ideas")	Asking Questions	5
	Talking to Learn	6
	Volunteering	7
	Answering When Called On	8

Table 2

LSP Assessment Instruments

Instrument Title	Characteristics Measured
Children' School Questionnaire (Phillips, 1966)	School anxiety and coping style
LSP Achievement Test	Test of knowledge of LSP concepts
Reading Test--Forms F-G Metropolitan Achievement Test (Durost et al., 1970)	Standardized reading achievement; primarily a test of comprehension
Word Meaning Test--Vocabulary subtest, Cognitive Abilities Test (Thorndike & Hagen, 1971)	Vocabulary; a measure of general mental ability (crystallized)
Matching Patterns--Standard Progressive Matrices (Raven, 1960)	Analytic reasoning: a measure of general mental ability (fluid)
Classroom Observation Instrument (Stallings & Kaskowitz, 1974)	Low-inference codings of teacher structuring, teacher demands for participation, and child participation
Teacher Questionnaire and Interview (PTE, 1978)	Questionnaire measured rapport, verbal fluency, anxiety, flexibility and classroom practices; interview measured attitude toward training and support for LSP

Table 3

Summary of Results

Outcome Measure	12-Cell Design	8-Cell Design
Treatment Effects		
January LSP Achievement	LSP F(1,20) = 11.66, p = .003 (1.16)[1]	Level F(3,24) = 7.76, p = .001 (2.27)
	LSP F(1,624) = 40.66, p = .001 (.50)	Level F(3,628) = 29.41, p = .001 (.58)
	DL F(1,624) = 5.43, p = .02	DL F(3,628) = 4.27, p = .005
May LSP Achievement	LSP F(1,20) = 4.11, p = .06 (.72)	Level F(3,628) = 16.07, p = .001 (.80)
	LSP F(1,624) = 10.94, p = .001 (.27)	
	DTL F(2,624) = 3.11, p = .05	
Vocabulary (covariate was Vocabulary pretest deviation score)	LSP F(1,20) = 5.86, p = .03 (.84)	Level F(1,627) = 3.52, p = .02 (.33)
	LSP F(1,623) = 7.42, p = .007 (.22)	
May Reading		Level F(3,627) = 2.50, p = .06 (.55)
Total Reading (covariate was standardized Total Reading pretest composite)	LSP F(1,20) = 4.80, p = .04 (.77)	Level F(3,627) = 2.13, p = .10 (.28)
	LSP F(1,623) = 3.40, p = .07 (.15)	
	TL F(2,623) = 6.11, p = .002	
	DTL F(2,623) = 6.29, p = .07	

ATI Effects

Outcome Measure	Predictor	df	F	b	R^2 Change
January LSP Achievement	(G x LSP)[a]	17	10.89**	.13	.24**
May LSP Achievement	G_d x LSP	17	1.53	.09	.04
	G x LSP		.89	-.03	.02
	G x A_x x LSP		2.69	-.10	.10**
Vocabulary	A_x x LSP	16	.21	.11	.01†
	(A_x x G_d x LSP)		4.42†	.35	.02
Raven's Matrices	G_d x LSP	15	2.88	.07	.02
	G x A_x x LSP		4.54*	-.08	.03*

†p < .10

*p < .05

**p < .01

[1] These numbers in parentheses reflect group differences in standard deviations (e.g. if $\bar{X}_{LSP}$ = 12 and $\bar{X}_{No\ LSP}$ = 10 and SD = 3.0, then difference = .67).

[a] Predictors in parentheses have 95 percent standard error confidence bands that span zero. This means that, despite the significant level of R^2 change, there cannot be confidence at the .05 level that the interaction effect is greater than zero.

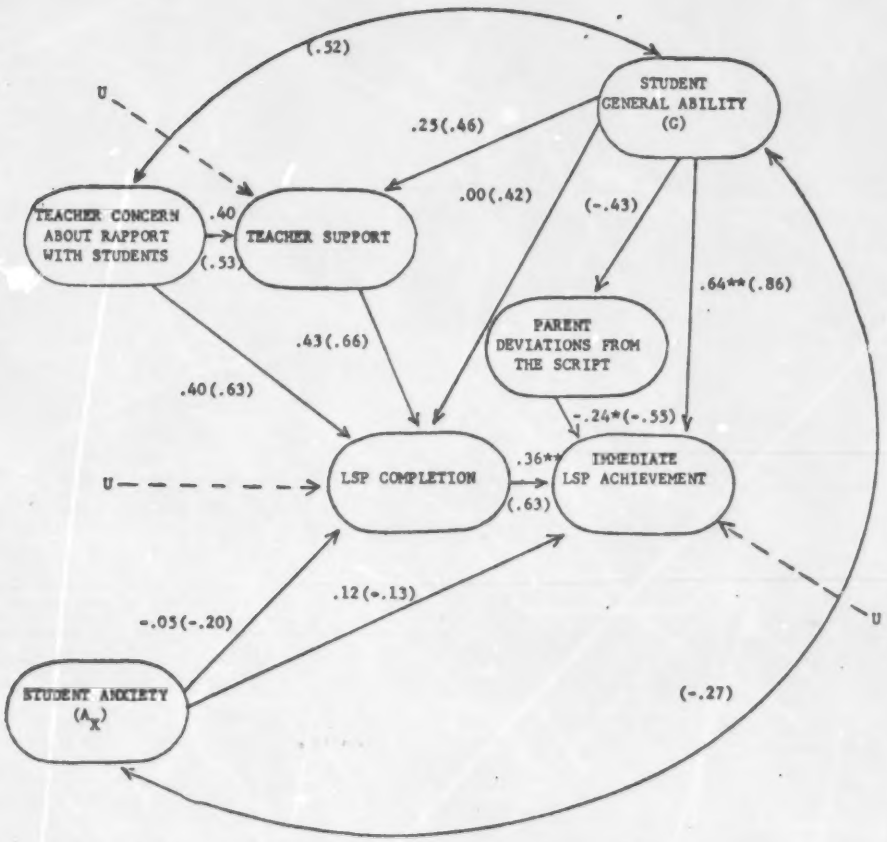


Figure 1. Reduced path diagram of LSP effects with January LSP Achievement. Numbers in parentheses are zero-order correlations. Other numbers are path coefficients. For $r \geq .41$, $p \leq .05$. *$p < .05$, **$p < .01$

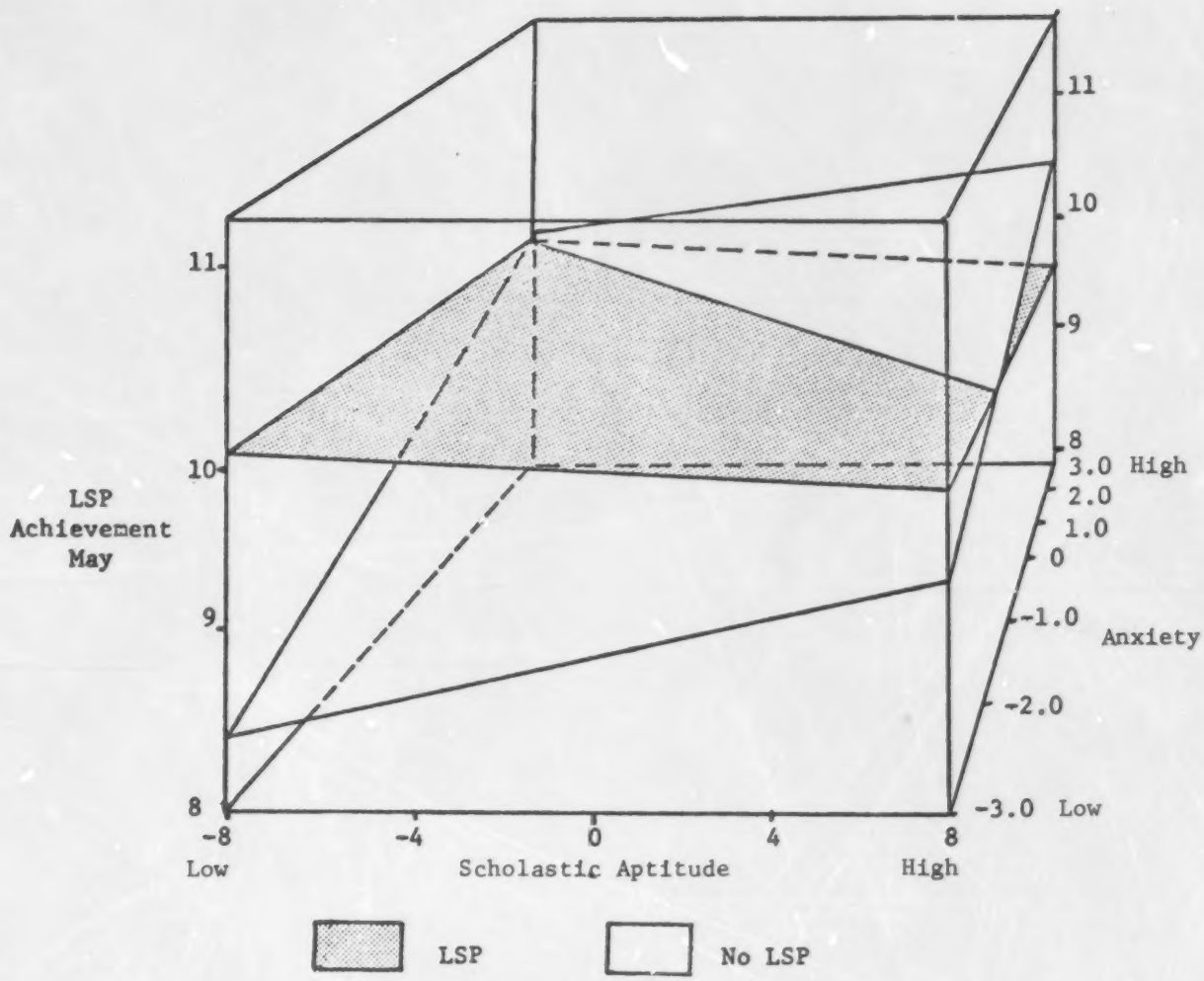


Figure 2. Regression planes for LSP-No LSP treatments using A_x and G as aptitudes and May LSP Achievement as outcome.

Appendix A

Sample LSP Exercise

LEARNING SKILLS PROGRAM

Trick Card #3

MEANING

These are the main things you should remember from a lesson. Often, important points are the things you will be tested on.

CLUES

This is very important for you to learn...
The main things you should remember are...
Don't forget that...
Try hard to get this...

GUIDE
EXERCISE 3

Goal: The child will identify important points in parts of different lessons.

1. INTRODUCTION

- Use Trick Chart to introduce important points and to review previously learned tricks. You will also need the Trick Cards for review and goals.

  Point to important points and tell your child:

> Today we will learn the third trick for making ideas orderly, important points. This is the goal for today.

  Now point to review and goals, saying:

> We have already learned the first two tricks, review and goals. Now that we know the goal for today's exercise, let's review what we've already learned.

  Have the child look briefly at review and goals Trick Cards. Read the meaning and the clues for each trick with your child. Now tell him:

> Important points are a little bit different from reviews and goals. Reviews and goals usually come at the beginning of a lesson, but important points can show up anytime in a lesson. Look now at your trick card for important points.

2. DEFINING THE SKILL

- Have the child read with you the meaning of important points on the card. Add the following:

> Teachers are careful to tell you if something you're studying is important, because they want you to do well in school. In order to do well on quizzes and to share your ideas in class, you need to know what is important to learn.

Have your child look at the Trick Card clues. Then tell him:

> Usually, the teacher will tell you what is important to remember by saying something like (read clues from card to child). Other things teachers and books do to tell you important points include: (1) saying something over and over again to make sure you "get it," (2) underlining things so the important parts stand out, (3) numbering important things, just like I'm doing right now, to help you keep the information orderly. As I said before, important points can come anywhere in a lesson. They can even be part of a review or a goal statement, for example.

- Give your child these examples of ways that teachers give important points:

> 1. It is very important for you to learn your multiplication tables.
>
> 2. The main things you should remember are how to write your N's and M's.
>
> 3. Don't forget (a) that the big hand tells the minutes, and (b) that the little hand tells the hours.
>
> 4. Try hard to get this now--the main idea of the story is that people are not perfect.

Now read your child these examples that do not have clues for important points.

> 5. What can you tell the class about Tom's story? (No clues to show this is an important point.)
>
> 6. In this lesson you will learn how to tie a slip knot. (States a goal, with no important points clues.)
>
> 7. The last lesson on fractions taught us how much 3/4's equals. (Review with no important points clues.)

3. APPLYING THE SKILL

- Read through the Trick Card clues once again.
- Read through Worksheet as directed.
- Have your child work through the longer section of the sheet as much as possible on his own. You may have to help with vocabulary words, such as message.

4. EXTENDING THE SKILL

Ask your child to watch his teacher this week to see if the teacher uses clues to tell important points. Give your child this summary:

> Of all the tricks for making ideas orderly, important points is probably the one that teachers use most often. Like in some of the examples we did, important points can be part of a review. And, goals by themselves are often important points, too. Important points don't have to come at the beginning or end of a lesson then, do they? Your teacher or a book can point out things for you to remember anytime in the lesson. Does your teacher do this much?
>
> If you are clever and try to spot the cues for important points, you will be able to keep your ideas together and then to do better in school. Remember, the clues for important points are (refer to Trick Card and read clues once again).

LEARNING SKILLS PROGRAM
MAKING IDEAS ORDERLY

NAME _____

WORKSHEET

Exercise 3

A. Read the example to child as if you were the teacher.

Question: Which lines have clues for important points? (Put a check by the line numbers.)

Answer: Line 3 and possibly 4. Lines 5-6 state a goal.

A.

1. Swimming is a sport that takes more smarts than strength.
2. When you learn to swim, you also learn two other important
3. things--first, to never swim alone, and second, to wait
4. awhile before swimming after you eat. When you are through
5. with all of your swimming lessons, you will be able to swim
6. with the best of them. You will know the backstroke, the
7. breast stroke, and the crawl.

B. Have your child read the example to you as a pupil in the class.

Question: Do any lines in the example contain clues to important points?

Answer: No. The example is all a review.

B.

1. Today my teacher told us more about the Indians of
2. America. She told us about the houses they build and the
3. way they can grow and catch food. She talked about
4. Indians for a long time. It will be much more fun to go to
5. the museum tomorrow and see the show they have on Indians
6. now that we know about them some. My teacher said that
7. seeing all the real things the Indians made should help
8. us remember what she told us in class.

C. Read the example to your child; again, you are the teacher.

Question: Which lines have clues to important points?

Answer: Line 5 has the clue "It is very important..." Point out that the important point clue comes inside a review.

C.

1. In our last art lesson we learned how to mix and blend
2. colors to make other colors. We learned that three key
3. colors can be used to make lots of other colors. These
4. three colors are called primary colors. They are red, blue,
5. and yellow. As I said last time, it is very important that
6. you remember these colors because you will have to use them
7. to make some other colors for me next week, and I won't be
8. telling you which colors to use. But now, who can tell me
9. what color comes from mixing blue and yellow?

D. Read these directions to child: This is a lesson from a school book. You should read it just as you would by yourself in class. Only, instead of answering the questions at the end of the lesson (you can, if you want to, but you don't have to), go over the lesson very carefully, and every time you see a clue for an important point, make a check next to the line number.

REMEMBER: IMPORTANT POINTS CAN COME ANYWHERE IN A LESSON--EVEN INSIDE REVIEWS AND AS PART OF GOAL STATEMENTS.

In this lesson, you will read a poem that has a very important message to tell. When you finish reading the poem, you should be able to say what this important message is. Read the poem now, and try hard to get the message before you answer the questions.

Fellowship

When a feller hasn't got a cent
And is feeling sad and blue
And the clouds hang thick and dark
And won't let the sunshine through,
It's a great thing, oh my brother,
For a feller just to lay
His hand upon your shoulder in a friendly sort of way.

It makes a man feel good inside
It makes the tear-drops start.
And you kind of feel a flutter
In the place inside your heart.
You can't look up and meet his eye,
You don't know what to say
When a hand is on your shoulder in a friendly sort of way.

Oh, this world's so full of wonder
With its sadness and its joys
Its winnings and its losses

Good for girls and good for boys.
And a good God must have made it,
At least that's what I say,
When a hand is on your shoulder, in a friendly sort of way.

Questions about the poem

1. There are a few key words circled in the poem. What one thing is the same about all of these words?
   a. They all describe colors.
   b. They all describe how the writer feels.
   c. They are all really long words.

2. Now look at the lines of the poem marked by arrows. What can you tell me about these three lines?
   a. They all talk about two men.
   b. They are shorter than the other lines in the poem.
   c. The lines all say the same thing.

3. What is the one main thing that this poem has to say?
   a. People should not be sad.
   b. It is sure nice to have friends when you are feeling sad.
   c. When you are down and out the only way is up.

The goal of this lesson was to teach you the main message of the poem--that "it is sure nice to have friends when you are feeling sad." You should have learned this message by going back over the poem, and looking at the circled words and the lines that are said more than once. The questions helped you do this. In our next poetry lesson, we will read about a boy who loves cars.

Answers: Important points are made in lines 1-4; 29; 34; 39; and 43-45. Point out that lines 43-45 also state a goal.
Correct answers to "lesson" questions are: 1. b; 2. c; 3. b;

An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade: The Observational Data

Nicholas Stayrook
SRI International

John Crawford
Center for Educational
Research at Stanford

Abstract

Low and high inference observational data that pertain to the teacher training intervention and to the confirmation of prior correlational results are reported. Analyses of variance, repeated measures ANOVAs, and analyses of covariance (with baseline rates as covariates) are presented for measures of program implementation. Although measurement problems exist for some variables, findings indicate that trained groups of teachers were behaving in accord with a number of the recommendations about effective teaching. Process-product correlations present evidence of confirmation of one half of the relevant results from prior studies. Implications and qualifications of the findings are presented, including suggestions to other researchers why may be considering this type of experiment in natural classrooms.

An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade: The Observational Data

During the 1976-77 school year, the Program on Teaching Effectiveness (PTE) of the Center for Educational Research at Stanford conducted a field-based experiment to test the effects of an inservice teacher training program designed to improve student reading achievement in the third grade. The other major purpose of the study was to intervene in randomly selected classes with a (home-based) parent-assisted instructional program designed to teach students a set of "learning skills" considered important for facilitating academic learning in the classroom. This paper reports only the observational data results related to the teacher training program. Corno (1978) presents and discusses findings for the parent-involvement program, including the use of observational measures in the assessment of aptitude-treatment interactions.

The recommendations in the teacher training program were derived from the results of four recent major correlational studies. The studies reviewed were: Stallings and Kaskowitz (1974); Brophy and Evertson (1974); McDonald and Elias (1976); and Soar (1973). These studies related specific teaching behaviors to student reading achievement and other outcomes in the primary grades.

Recommendations for the inservice teacher training program were developed by selecting variables from the four studies that showed significant or near-significant correlations with adjusted student reading achievement. Analysis of the four studies produced a list of 194 specific classroom behaviors or activities that were shown to be significantly related to student reading achievement. Of the 194 variables, 125 were deemed to have potential relevance

for recommendations to teachers. "Information sheets" reported in Volume I of the final report on the experiment (see PTE, 1977) were created for each of the 125 variables. Sixty-nine of the 194 variables were not included in the set of information sheets either because they overlapped with variables already used, or because the definition of the variables was too vague to permit specific recommendations to teachers.

The 125 variables included in the information sheets were grouped into three broad categories: Behavior Management, General Instructional Methods, and Questioning and Feedback Techniques. The information sheets for the variables included: (1) the operational definition of the observational variable, (2) its correlation with adjusted student reading achievement, (3) its average frequency of occurrence and standard deviation, and (4) an interpretation of the meaning of the process-product relationship--as given by the original authors.

Eighty-one of the variables were actually used in the derivation of training recommendations. The reason for the reduced number of variables was that some redundancy still existed within the set of 125 variables, or that there was difficulty in telling teachers how (and how much) to use some of the behaviors. For example, many "high inference" process variables and factor score composites were not used because they yielded little specific information about behaviors the teacher should be performing in his or her classroom.

This paper addresses two major issues through analysis of the observational data. The first issue is whether the teacher training program conducted in November and December of 1976 produced significant differences in teaching behaviors between the groups of teachers who received the training and the

group of teachers who were only observed. The second issue to be addressed is whether the behaviors that were significantly related to student reading achievement in the four prior correlational studies were also significantly correlated with the measures of student reading achievement in the present study.

Measurement of Behaviors

The operational definitions of relevant process variables were used in modifying the SRI Classroom Observation Instrument (COI) that was used in evaluating Project Follow Through by Stallings and Kaskowitz (1974). The COI is a general time-sampling observation instrument with which the observer records specific behaviors by indicating who is performing the behavior, to whom the behavior is directed, what the behavior is, and how that behavior is being performed.

The 81 operationally-defined teaching behaviors were "mapped onto" the COI. This process of mapping the variables from one observation system (e.g., the Classroom Dyadic Interaction System used by Brophy and Evertson) onto the COI proved to be very difficult. Some of the difficulty lay in the fact that definitions of the observational variables were not always clear. Another difficulty was due to the fact that the COI is a time-sampling instrument, while the system used by Brophy and Evertson employed continuous recordings. We believe that these logistical differences between different coding systems have an important impact on the kinds of information resulting from the systems.

Another major problem was that the creation of the actual variables of interest from the modified COI was a post-hoc procedure. Because the SRI system is so general, coders were trained only in the definitions of each discrete

"bubble" (code) to be marked; they were not trained to make codes that represented lengthy sequences of interactions, as were coders using the Brophy-Evertson system. It is possible to create a tally "key" composed of a string of "bubble" codes from the COI (sometimes having to span three or more separate frames of coding) that nominally represents the same phenomenon that was coded with a standard code in the Brophy-Evertson system, but the true equivalence of these measures has yet to be established. In fact, reliability studies of the process variables suggest higher reliability for variables originating in the Stallings and Kaskowitz (1974) study, indicating that variables from other studies are not always represented adequately with the COI.

Table 1 presents the list of observation variables grouped by the Behavior Management, Instructional Methods, and Questioning and Feedback categories. For each major category, the specific variables within the category were grouped according to the study from which they came. In addition, the instruments used to measure the behaviors are indicated. In the present study, four instruments were used to measure the classroom behaviors of our sample of teachers. These four instruments were (1) the Five-Minute Observation (FMO) from the COI, (2) the high inference ratings, (3) the Classroom Checklist from the COI, and (4) the curriculum survey.

Insert Table 1 about here

The FMO is the basic time-sampling unit in the SRI Classroom Observation Instrument. Every 15 minutes throughout the school day the observers were

instructed to record for five minutes the discrete behaviors occurring during that period. During the same 15-minute interval, the observers were instructed to fill out the Classroom Checklist. The Classroom Checklist indicates what activities are occurring in the classroom, the grouping patterns being used in those activities, and whether or not the teacher (or an aide or volunteer) is associated with a particular group of students.

High-inference ratings were made at the end of each morning and each afternoon in each classroom during the posttraining phase between January and May. Observers were asked to rate on a five-point scale many behaviors that were not measurable with the low-inference SRI system. The curriculum survey, a 29-item questionnaire dealing with some additional teaching behaviors and with the class's curriculum "content coverage," was administered to each teacher in May of 1977.

Table 1 lists 64 of the observation variables. The major reason for this reduction from the 81 variables cross-referenced in the teacher training materials is that analysis of the observational data showed 17 of these behaviors to be very low-frequency (or zero-frequency) behaviors as coded with the COI in this sample of classes. The variables listed in Table 1 are those variables on which observation data were obtained, and concerning which the teacher training materials provided recommendations. Generalizability analyses of reliability across observers suggested that the variance component associated with the teacher/class facet (in a teacher by observer design) was insufficient for trustworthy analyses of 13 of the FMO variables. These 13 variables are listed separately at the end of Table 1. Eight of the 13 are Brophy-Evertson variables, four are Stallings-Kaskowitz variables, and one is from the McDonald study. Readers interested in a complete description of the generalizability analyses should see

Volume III of the Final report, Appendix D (PTE, 1978). These 13 variables were not entered into further analyses.

Observation Procedure

The selection and initial training of observers were performed cooperatively by the PTE staff and the SRI staff. Raters were trained for a period of two weeks prior to any data collection. Agreement between observers was estimated at the beginning of the observation in September 1976 and again in January 1977. The refresher training session in January 1977 lasted about one week.

The average "percentage agreement" between pairs of observers was computed during the refresher session in January 1977. Each observer completed 12 FMOs over a three-day period; all observers simultaneously coded the same teacher (a PTE consultant, not one of the sample teachers) for each FMO. Table 2 reports the average percentage agreement between observers during that reliability check for each of the specific codes ("bubbles") for the FMO portion of the instrument. At the same time, observers were also asked to fill out the high-inference observations so that agreement could be checked on those measures. Table 3 presents the average interrater agreement for eight treatment-related high-inference observation variables.

Insert Tables 2 and 3 about here

The schedule of observations was influenced by the design of the study and by a desire to observe classes on a sufficient number of occasions for the assessment of treatment effects on implementation. The study progressed through three major phases: (1) baseline observations started around mid-September and

extended through the end of October; (2) training phase observations began the first of November and continued through mid-December; and (3) posttraining observations were taken between January and mid-May.

During the baseline phase each teacher was observed on four full days to obtain an estimate of the degree to which the teacher may have been performing the recommended teaching behaviors prior to the training.

Following baseline observations, the teachers were assigned to treatment groups by means of a stratified random assignment procedure, stratifying on mean class achievement. The three teacher training treatment groups were: "maximal" training, "minimal" training, and observed-only (within each of these groups, stratifying and random assignment was used again in assigning classes to the Learning Skills Program groups). During the training phase of the study the maximal training group received the written teacher training materials and a two-hour workshop with PTE staff once a week for five weeks. The minimal training group received the written teacher training materials only. The control group did not receive any training, but were observed. A more complete description of the training is given by Crawford and Stallings (1978). During the training phase of the study most trained teachers received five and control-group classes typically received four full days of observation.

During the posttraining phase of this study, each teacher was observed on at least seven days, at the rate of about once every two weeks. Observations made during the posttraining phase were used along with training phase data as the major criteria for assessing treatment implementation.

Data Analysis Procedure

The low-inference COI data were collected in a machine-scannable booklet.

Each booklet recorded four FMOs and represented one hour of classroom observation (when the FMO portions were not being filled out, coders were completing the Classroom Checklist or were recording context-identifying information). The booklets were checked for accuracy and sent to Intran Corporation (Minneapolis, MN) which optically scanned the booklets and recorded the data on a computer tape. A total of 8,503 FMOs were taken in the 33 participating classrooms during the course of the school year.

The observational data tapes were checked and cleaned, and then submitted to a scoring program for tallying the frequency of each observational variable of interest. The Classroom Checklist data were also prepared in a similar manner. The high-inference ratings and the curriculum survey were not optically scanned, but were coded and readied for analysis by PTE staff.

All FMO variables and Classroom Checklist variables were next aggregated (arithmetically averaged) over five-minute observations within phases of the study. For some implementation analyses, no distinction was made between training and posttraining observations. Each FMO consisted of about 60 frames of coding, with one frame being recorded approximately every five seconds. Since there was some variation in the number of frames per FMO, the data were standardized--made comparable from one FMO to the next--by dividing frequency of occurrence for each variable by the number of frames making up that FMO. Therefore, the original metric for the FMO variables was "rate per frame." If multiplied by 60 (the modal number of frames per FMO) one would get "rate per five minutes;" multiplying by 720 yields "rate per hour," etc. For inferential analyses, proportions (as in Brophy and Evertson, 1974) and standardized Z-scores ($\bar{X} = 0.0$, SD = 1.0) were created--in order to permit the creation of composite scores and

to make assessment of effect sizes (percentage of an SD) simpler. The FMOs were aggregated over the whole year of observations for a given teacher for the process-product replication of prior studies. The high-inference ratings were made only during the posttraining phase of the study; hence, posttraining only observations were aggregated to obtain a score on each high-inference variable for each teacher. The curriculum survey was filled out by each teacher during the last month of the school year. Scores on curriculum-survey items complete the posttraining implementation data set.

The discussion of results is divided into several parts. The first part gives the results pertaining to the degree of treatment implementation, reporting analyses of variance, analyses of covariance with baseline rates as covariates, and repeated measures analyses. The last section reports results related to the confirmation or disconfirmation of the correlational findings from the prior studies.

Assessment of Treatment Implementation

The proper analysis of observational data for determining degree of treatment implementation in this study cannot be done on the basis of already available formulations and procedures. Upon first consideration, one might suppose that between-group analyses of variance of a measure of conformity to training recommendations would yield a straightforward and obvious answer to the question "Were the trained groups of teachers behaving in accord with the training principles in the inservice program?" We will present such analyses, but will also attempt to demonstrate that this approach ignores many of the inherent complexities in this study; these kinds of group comparisons may eventually prove to be more useful for studies of rates of behavior pertaining to independently developed

and administered programs (with different training content across programs), such as in the analysis of implementation of several different Follow Through Models of instruction (see Stallings and Kaskowitz, 1974).

An example using one of the recommendations from the General Instructional Methods Training packet may illustrate some of the difficulties. Based on a number of correlations at the third grade level in the Stallings and Kaskowitz (1974) report, training group teachers were advised to make substantial use of large group (defined as more than 8 students) instruction. This was interpreted to teachers as a potentially efficient way of providing a maximum amount of instruction to the largest number of students because whenever the teacher is instructing one or two students at a time, the other 20 to 25 are on their own. The latter method may not be the most effective method of teaching in the primary grades.

However, the recommendation was not that large group instruction should occur 100% of the time. Interpolation using means and standard deviations from the Stallings and Kaskowitz report yielded a recommendation that teachers should teach to large groups "at least one-third to one-half of the time." This recommendation was derived by adding approximately one standard deviation to the mean of the process measure for use of large group instruction in the prior study. If the variable was normally distributed in the previous study, then about 34% of the classes had naturally occurring rates between the mean and one standard deviation above the mean. The major point to be made is that, in the present study, the one-third of the teachers not given any training could relatively easily (depending on the curriculum being used, the number of aides in the room, etc.) be using large group instruction one-third to one-half of the time. If this were the case, then the between-group analysis of variance might yield

an answer of "no implementation" for this measure, while, in fact, the training group teachers could be in perfect conformity with the training recommendation --all groups could be using large group instruction 30-40% of the time.

This same problem, or analogous problems, exists for almost all of the process/implementation variables, primarily because the recommendations were derived from naturalistic observational studies. The behaviors had to have been found occurring with some natural variation, in some sample of classes, before the behaviors could have been incorporated into the teacher training program. Between-group ANOVAs might be more appropriate if, for example, the treatment consisted of experimenter-developed curricula and instructional procedures based on cognitive development theory. A treatment such as this would not be so likely to be occurring in no-intervention classes, and ANOVA of implementation indices should be an accurate reflection of degree of use in experimental group classes.

Nonetheless, an argument can be made that, even for the "naturally occurring" treatments of the type employed here, with many observations taken over several months, some overall group differences should be expected to emerge. That is, we would hypothesize that even though many treatment-related behaviors do occur in control group classes, the trained teachers should be using the recommended strategies more often and/or more consistently. Analyses of variance suggest the degree to which this hypothesis was supported, and also provide valuable descriptive information about the randomly assigned groups of teachers.

The design of the study also has an impact on the "statistical conclusion" validity of analysis of implementation. The training program consisted of more than 20 different recommendations about ways of teaching thought to be effective.

In order to approximate the conditions under which an inservice program would typically be presented, all teachers assigned to training groups received the entire program. That is, the discrete behaviors were not manipulated independently of each other as they might have been in a more complex factorial or fractional factorial design. Strictly speaking, then, the analysis of individual components of the program should not be conducted as if the behaviors were independent of each other--it is certainly possible that high or low implementation on one variable could lead to (causally influence) high or low implementation on other, related variables.

However, following this line of reasoning to its logical conclusion leaves us with the possibility that the only "legitimate" implementation analysis would be one in which a summary, total implementation score is entered into inferential analysis. In our judgment, such statistical considerations should not override the need to present the maximum amount of information. Given an observable treatment effect on a summary total implementation score (as is reported by Crawford and Stallings, (1978) the first question that comes to mind is "What specific behaviors are most responsible for the overall effect?" Most of this paper is concerned with presenting possible answers to that question through analysis of discrete behaviors from several different perspectives.

The first section of the implementation analyses gives ANOVA results for all post-baseline observations (training and posttraining phases aggregated together). For two major reasons, it was considered appropriate to combine training and posttraining data. First, the generalizability analyses briefly mentioned earlier suggested that some process measures may have less-than-desirable reliability. The more data (observations) a particular measure

is based on, the higher its reliability--in the same way the reliability of a pencil and paper test score is increased by increasing the number of items. So, the reliability of training and posttraining data aggregated together will be higher than the reliability of posttraining-only data. Secondly, we could argue psychologically that the training phase data should comprise part of the basis for the dependent variables because the teachers were likely to be implementing while the training was occurring. Further support for the potential impact of the training phase data was demonstrated by the treatment effect on mid-year reading comprehension reported in Volume II of the Final Report on the experiment. Those results indicated that, as soon as 6 weeks after the completion of the major portion of the training (one "refresher" training packet was sent to teachers during the spring "posttraining" phase), the adjusted achievement in the trained classes was higher than in untrained teachers' classes. Since the groups were equivalent in achievement at the beginning of the year, it is reasonable to assume that the experimental group teachers were implementing teaching strategies between November and January that account for their students' enhanced mid-year achievement.

The section of results and discussion following ANOVA of post-baseline measures focuses on analyses of covariance of low-inference variables for which corresponding scores exist for baseline and post-baseline data. These include the FMO and Classroom Checklist variables from the COI. Although ANCOVA is not a common procedure for analysis of process variables, there is reason to expect that the evaluation of group differences after baseline variation has been removed will yield relatively unbiased estimates of differences between groups. Homogeneity of regression tests (in effect, aptitude-treatment interaction analyses) are also reported.

The last major group of analyses reports the repeated measures analyses of variance in which the "trials" factor is defined by "time of year"--i.e., baseline,

training, and posttraining phases. These analyses, again only for the FMO and Classroom Checklist variables measured regularly between September and May, provide several interesting sources of information about changes across time. The most interpretable effect would be a treatment by time-of-year interaction, in which the three treatment groups began the year with equivalent rates of behavior, but then diverged during training and posttraining phases, with trained teachers' classes changing to be more conforming to the recommendations, but the control classes maintaining an approximation of their baseline rate throughout the whole year. Although the data do not uniformly conform to this idealized pattern, the results suggest that this type of multivariate method can provide many interesting clues about the influences on classroom processes.

The last set of results on implementation effects concerns a number of variables that had shown differential patterns of process-product correlations in the Brophy-Evertson (1974) study according to whether the classroom was lower SES (lower to lower-middle) or higher SES (middle-middle to upper-middle). Behavioral typologies of lower and higher SES students were developed from analysis of a technical report giving process differences by SES level. Then, whenever the training recommendations dealt with the SES-differentiated behaviors, teachers were advised to teach their students differently according to the child's characteristic "academic orientation." "Academic orientation" (low vs. high) was used as a within-class proxy for SES, since it was thought that the behavioral responses of the students were more likely to correspond to the attributes actually responsible for certain teaching methods being more effective with certain types of students. Interested readers should see the first and third volumes of the final report for a more complete discussion of the

background and rationale for this part of the study. The second column of Table 1 in this paper indicates which variables were SES-differentiated.

The results for confirmation or lack of confirmation of prior correlational findings are the last data presented in this paper. In these analyses, data from the whole year were aggregated to obtain the best estimate of the relative occurrence of the treatment-related process variables. Then, these process measures were correlated with end of year achievement that had been adjusted to remove pretest and district differences. The discussion of the process-product matrix and the agreement with prior studies refers to differences in samples and problems of measurement and analysis.

Analysis of Variance of Training Components

As mentioned, these ANOVAs are based on data aggregated over training plus posttraining phases. Table 4 shows how the variables were combined to give measures of training components within the Behavior Management, General Instructional Methods, and Questioning and Feedback categories. The variables in Table 1 are somewhat redundant, particularly in the General Instructional Methods category. This resulted in composite scores representing some of the more "molar" recommendations, such as "teachers should decrease the number of non-academic interactions." Variables without useable class-related variance were excluded from this analysis, as were the SES differentiated findings (with one exception--the variable "percentage wrong answers", a measure related to the difficulty level of the teacher's questions, was included so as to have a better represented Questioning and Feedback category--this measure showed strong evidence of behaving in this high SES sample as a "main effect" or undifferentiated variable).

Insert Table 4 about here

Table 5 summarizes the treatment main effects analysis of variance for the 16 training components listed in Table 4. The ANOVA design was District by Treatment, so all the F-ratios in Table 5 have 2 and 27 df. The direction of the means is indicated; the actual means and standard deviations are presented in Appendix A at the end of the paper.

Insert Table 5 about here

The results in Table 5 show that about half of the training components evidence successful training implementation in this between-groups analysis. Within the Behavior Management category, trained teachers were observed using a system of rules for minimizing interruptions more often than controls. Minimal and maximal training teachers also more successfully avoided "timing" errors (allowing misbehaviors to continue and spread), and were rated higher on "Withitness"--an index of awareness of behavior problems and communication of that awareness to students (see Kounin, 1970).

The components in the General Instructional Methods category showed only one significant between-group difference, for the variable representing minimization of direction giving, organizing, and transition activities. Even though the other variables in this category were not significant, three of the other five did have means for the control group lower than the trained groups (this, combined with the large differences for the measure of direction giving, produced an overall effect for General Instructional Methods--see Crawford and Stallings, 1978).

Three of the four Questioning and Feedback variables showed significant or near-significant differences in the right direction (control group lowest). The two variables measuring methods for selecting students for response opportunities showed substantial differences favoring the maximal, then the

minimal, and then the control group. Teachers were advised in the training program that: (1) pre-selecting students before asking the question or calling on non-volunteers was a good way to maintain an equitable distribution of questions to all students in the class (and thereby encourage a maximum amount of involvement in the discussion), and (2) allowing student "call outs" during questioning sessions would only contribute to a disruptive atmosphere and insure that a select group of students would be responding. The data suggest they were able to follow this advice effectively.

Overall, the minimal training group had a higher implementation mean score than the control group on 13 of the 16 components. The maximally trained group had a higher mean than the control group on 11 of the 16 components. And the minimal group was higher than the maximal group on 12 of the 16 components. The evidence is good that the trained groups excelled the control group on the criteria of implementation and also that the minimal training group excelled the maximal training group on the criteria.

In general, these findings look promising. Considering the earlier discussed caveats about reasons for not establishing between-group effects, the occurrence of significant group differences in the right direction for 7 of 16 components indicates that the inservice training resulted in many detectable differences with trained groups of teachers conforming to the recommendations. The apparent superior implementation for some of the components by the minimal training group--in comparison to the maximal group--was not expected. Perhaps the mediating influence of certain teacher characteristics produced this finding. Crawford and Stallings (1978) discuss the teacher characteristics such as verbal ability and initial "structuredness" of teaching style as related to implementation.

Analysis of Covariance of Implementation Measures
(Baseline Rates are Adjusting Covariates)

A natural question about implementation effects is "What are the group differences when statistical controls are employed to remove variation associated with rates of occurrence at baseline, prior to intervention?" As such, this is an analysis of covariance question, similar to analysis of posttest achievement with pretest achievement as a covariate.

Analysis of covariance is a complex procedure, and when used correctly, it can be a powerful statistical technique. It requires more assumptions than does ANOVA, and the violation of those assumptions can influence considerably the answers one obtains from applying ANCOVA procedures. Many statistical articles and chapters (see Kerlinger and Pedhazer, 1973, or Ward and Jennings, 1973) have discussed at length the assumptions for ANCOVA. Accordingly, we will be brief in discussing the underlying assumptions here, with particular attention to the degree to which our data meet the requirements.

The most basic assumption of ANCOVA (over and above independence of observations, homogeneity of variance, and normality of distribution which should be considered for ANOVA or ANCOVA) is commonly referred to as the "linearity" assumption. Stated briefly, this assumption requires that for each unit change in the covariate there is a constant difference in the dependent variable. If the covariate and the dependent variable are not correlated, then the regression line representing the pre/post relationship will be horizontal and indicate no difference in the dependent variable as the covariate changes. If the pre/post scatterplot is more spherical (as in $r_{xy} = .15$) than elliptical, then the dependence of the outcome measure on the covariate is almost non-existent. Therefore, only the low inference variables with statistically significant ($p \leq .05$) correlations between baseline and post-baseline observations were selected for further analysis leading to ANCOVA.

In addition to having to meet the linearity assumption, the data must also demonstrate homogeneity of regression coefficients across the treatment groups whose adjusted means (i.e., regression-computed intercepts) will be compared in the inferential tests. This is an aptitude-treatment interaction (ATI) question concerning whether the pre/post relationship varies with treatment group. In this study, an interpretable ATI would be found if the control group's slope (training and posttraining implementation regressed on baseline) were steeper than the trained groups' slopes, suggesting that in no-intervention classes, the post-baseline rates were heavily contingent on baseline rates, while the relationship for the group of trained teachers was decreased, presumably because of the intervention. This kind of ATI is similar to what Salomon (1972) has termed "compensatory" ATI, meaning that the treatment has the potential for mediating or compensating for the pre-treatment variation. In our case, the treatment might lessen the impact of baseline rates of occurrence on post-baseline rates. In discussing the results of the homogeneity of regression tests, the only ATI we will take seriously are those in which the control groups' slopes are higher than the experimental groups' slopes.

Tests of slope differences were made by comparing the predictive accuracy of two linear models for each variable: one model that "allows" separate slopes and intercepts for the 6 treatment-by-district groups, and another model that permits different intercepts, but has only one (pooled slope) parameter for estimating the pre/post relationship. A significant F-ratio would reflect a substantially higher residual sum of squares in the latter ("restricted") model and suggest that the use of a pooled slope (as in ANCOVA) is not justified. A non-significant F-ratio would be a signal to continue with ANCOVA using the pooled slope, as it is statistically a reasonable pre/post relationship estimator for all treatment groups

Table 6 below presents the relevant regression statistics for the (ATI) homogeneity of regression tests. Variables not appearing in Table 6 but in Table 1 do not have significant pre/post correlations, or are not reliable, or are SES-differentiated.

Insert Table 6 about here

The regression analyses summarized in Table 6 do not indicate that more than a chance number of variables have different slopes for the different treatment groups. However, this result may be due in part to the small number of subjects, and hence, unreliable estimation of parameters (and also a small denominator df for the F-ratio). Of the five variables that showed a tendency for heterogeneity of slopes, only one, QF02--Student "Call outs," conformed to the meaningful pattern of the control group having the highest slope in both districts. Three of these five variables had control group slopes higher than treatment slopes in one district, but not the other. In general, there is little evidence suggesting that the pooled slope is not an adequate regression line for use in ANCOVA. Table 7 below summarizes the ANCOVA results for the 16 variables from Table 6 that showed no trend toward heterogeneity of regression.

Insert Table 7 about here

The results in Table 7 do not suggest any different conclusions from those obtained by analysis of variance of unadjusted scores. Three fourths of the variables entering into the analyses of covariance are from the Instructional Methods category. Therefore, the lack of between group differences on these adjusted scores (ANCOVA is computationally the same as ANOVA on adjusted scores) agrees with the earlier reported findings that components of the Instructional Methods training do not show large mean differences between treatment groups. The hypothesis that such differences would appear when baseline rates were

controlled for does not appear to be supported. Also, with the possible exception of a few variables, little evidence of ATI was found--i.e., for the measures analyzed in this fashion, the treatments do not seem to have mediated substantially the relationship between baseline and post-baseline rates of behavior.

The lack of group differences for the discrete components of the General Instructional Methods categories is probably due to the nature of those recommendations. Those strategies were either very global ("try to decrease non-academic interactions," "teachers should increase the amount of student engagement with the academic material," etc.) or were dealing with aspects of teaching over which the teacher may not have complete control (e.g., grouping patterns, which, in this study, were influenced by most teachers' involvement in state-funded "individualization" programs). As noted in the example about use of large group instruction, these are also the behaviors that control group teachers would be most likely to be naturally implementing at about the same level as the treatment groups. The means in Table 7 (when extrapolated to a "rate per hour" metric) for the FMO variable measuring "adult attention to large group" do, in fact, show that all groups were within the recommended range of "1/3 to 1/2 of the time"--the maximal and minimal training groups instructed large groups between 25 and 26 minutes per hour, on the average, and the control group averaged slightly less than 25 minutes per hour. This indicates that the trained groups were conforming to the recommendation about large group instruction, and <u>also</u> that they probably would have been doing so even without the treatment

intervention.[1] If one is interested in establishing between-group implementation effects, the best bets are probably those behaviors that control group teachers would be unlikely to be exhibiting naturally and consistently.

Repeated Measures ANOVAs

The analyses reported thus far have used as dependent variables the implementation indices based on training and posttraining data aggregated together. The next set of results are based on a repeated measures analysis in which the "trials"--i.e., the levels of the repeated measures factor--are based on the "time of year." "Time of year" in this study refers to the three major phases of the experiment: Baseline, Training, and Posttraining. For these repeated measures ANOVAs, the data were aggregated separately within phases. Then the three scores for each implementation variable were entered as the "Y" (outcome) matrix in the multivariate repeated measures analyses. The major reasons for doing this analysis were to examine the data for treatment by trials interactions (to see if the groups' rates change differentially over time), and also to determine the degree to which "time of year" may evidence main effects. It could also be argued that a treatment main effect without a treatment by time-of-year interaction should not be taken seriously if the control group means at baseline were in the same relative position (compared to the other groups) as in the post-baseline phases. Table 8 below summarizes the repeated measures ANOVAs, giving F-ratios and means for low inference FMO variables.

Insert Table 8 about here

---

[1] Part of the problem here may lie in the rationale used in deriving "optimal rates." If a process variable was found to be positively correlated with achievement in the prior study, then the experimental groups in this investigation were advised to perform the behavior at a rate that was about 1 standard deviation above the mean (using the mean and standard deviation from the previous study). It is possible that the $\bar{X}$+1SD in the prior study's sample of classes is roughly equivalent to the (naturally occurring) sample mean in this study. This appears to be the case at least for the "large group instruction" variable.

The first variable in Table 8, a measure of preselecting or calling on non-volunteers, showed a significant treatment effect and a time-of-year effect, but no treatment by time-of-year interaction. The means show that the trained groups were higher than the control group in each of the phases, including baseline. This analysis indicates that the effect reported for this variable earlier (in Table 5) should be viewed cautiously, since the experimental groups were higher than the control group prior to intervention. The repeated measures analysis also shows these means generally increasing from baseline to training to posttraining, regardless of treatment group.

Analysis of the next variable, a measure of student "call outs", shows a main effect for treatment and a treatment by time-of-year interaction. The recommendation to teachers for this variable was that they should decrease the amount of call outs so as to avoid disruption of the discussion. The means in Table 8 indicate that the training groups did decrease the amount of call outs in their classrooms from baseline to training and posttraining, while the control group increased their rate from baseline to the post-baseline phases. At baseline, the control group mean was in the middle (higher than the maximal group but lower than the minimal group).

The next variable is a measure of the proportion of student responses coded as correct. This variable shows a treatment main effect and a time-of-year main effect, but no treatment by time-of-year interaction. The order of the means at baseline were: minimal group highest, then maximal, then control. During training and posttraining, the order of the means remains the same. The time of year main effect was due to means for all groups being lower during the training phase than during baseline and posttraining.

The next variable is a measure of the proportion of student responses coded as incorrect. This variable is interpreted as a "proxy" measure of the difficulty level of the teachers' questions. This measure showed main effects for treatment and time-of-year, and also a tendency for a treatment by time-of-year interaction. For high SES or "more academically oriented" students, teachers were advised to increase the difficulty level of their questions. The treatment by time-of-year interaction seems due to the maximal and minimal training group means increasing to a larger degree than the control group mean increase from baseline to training and posttraining. If the training and posttraining data are averaged, the maximal and minimal group increases are 1/3 to 1/2 greater than the control group's increase. All groups' scores tended to increase from baseline to posttraining, except for a slight decrease in the maximal group from baseline to training, followed by a large increase from training to posttraining.

The next dependent variable in the table that shows some significant or near significant F-ratios is teacher criticism following student incorrect answer. The only source that was significant for this variable was the time-of-year main effect, such that the overall means were higher for posttraining than for baseline and training.

The next two variables in Table 8--teacher rephrases or gives clues and teacher repeats questions--show no treatment by time-of-year interactions, but do yield evidence of treatment main effects and time-of-year main effects. The direction of the means for these variables suggests that teachers in trained groups were not conforming to the recommendations, as the "rephrases or gives clues" variable was supposed to be used primarily with lower SES or less academically oriented students, and the "repeats question" variable should have been minimized for all students. The control group means are the lowest means during training and posttraining for both of these measures.

The next three variables relate to rates of self and opinion questions (non-academically related questions). One of these measures, the ratio of self questions to academic questions, shows a significant treatment by time-of-year interaction in the direction that would be expected, given the training recommendations. The maximal and minimal group means decreased from baseline to post baseline phases, while the control group means increased from baseline to training and then decreased from training to posttraining--such that baseline and posttraining means were approximately the same. The time-of-year main effect for this measure seems primarily caused by the maximal and minimal groups' decrease in amount of non-academic questioning after the beginning of the training program. The treatment main effect on the variable measuring ratio of opinion questions to academic questions cannot be attributed to the intervention, since the control group was lower than the trained groups at baseline, as well as during the post-baseline phases.

Following the self and opinion questions variables is a measure of teacher use of sustaining feedback. This variable shows treatment main effects, favoring the trained groups, and a time-of-year main effect such that all groups tended to increase from baseline to training to posttraining. Effects for this variable are difficult to interpret because of problems in measuring this process with the low inference system used in this study (and also because the feedback variables were typically SES differentiated, and therefore not "main effect" variables).

"Supervising" and "instruction", the next two variables in Table 8, both show relatively strong main effects, and the supervising measure shows a tendency for a treatment by time-of-year interaction. However, the recommendation to teachers was that they should decrease the amount of supervising. The direction of the means shows an increase for the maximal and minimal groups from baseline

to training and posttraining, and a slight decrease only for the control group from baseline to training, and then an increase from training to posttraining. The low inference data for this measure do not support the idea that experimental group teachers were able to decrease the amount of supervising after the training intervention. The variable measuring teacher instruction shows a decrease for all groups from baseline to training and posttraining. This may reflect a general tendency in schools for more academic activities to occur earlier in the year.

The measure of teacher non-academic commands, requests, and questions shows significant time-of-year and treatment by time-of-year F-ratios. The means for this variable decrease consistently for the trained groups, but increase for the control group from baseline to training, and then show a decrease to posttraining. From these analyses it appears that the trained groups of teachers were able to effectively decrease the amount of non-academic interactions while the untrained groups of teachers only decreased from training to posttraining, but showed a large increase from baseline to the training phase.

The measure reflecting adult academic commands, requests, and questions directed to groups of children shows no treatment main effect or treatment by time-of-year interaction, but a near-significant time-of-year main effect was established such that the overall rates are lower for the later phases. Although the interaction was not significant this trend is more consistent for the trained groups of teachers than for the control group, who increased from baseline to training. The next variable shows approximately the same pattern as the prior one (the only difference between the two is that the latter does not restrict the teacher's actions to pertain to groups of children). There is an overall decrease from baseline to training to posttraining. Again, this may reflect the naturally occurring situation of less academic activities going on later in the year, when compared to early in the year.

The next variable, a measure of non-academic child responses, shows a time-of-year main effect and a treatment by time-of-year interaction. The minimal group means decrease, as they should according to the recommendations, from baseline to training and posttraining. However, the maximal group means increased from training to posttraining, after decreasing from baseline to training. The control group means were in an opposite pattern from the maximal group means, with the training phase mean being higher than baseline and posttraining means.

The variable measuring amount of academically-related adult instruction shows consistent main effects for time-of-year and no other significant or near significant sources. The later the time of year, the fewer the amount of academically-related adult instruction interactions.

The measure of child to adult interactions (all verbal interactions except child responses) indicates that only the maximal training group was able to decrease and maintain a decrease from baseline to the post-baseline phases. The minimal group means on this variable did not change, and the control group means decreased to training, but then increased at posttraining.

The next two variables are measures of teacher attention or instruction with a large group. Both of these variables show time-of-year main effects and treatment by time-of-year interactions. The first one of these two shows decreases from baseline to post-baseline phases regardless of treatment group. The treatment by time-of-year interaction is apparently caused by much greater reduction in the control group. Since this variable was positively correlated in the prior studies, the recommendations advised teachers to use this type of instruction rather than to try to avoid it. Apparently, all three groups changed in the nonrecommended direction, but the control group did so to a greater degree than the trained groups. The next variable shows a slight increase from baseline to training for the control group. The minimal group mean increased from training to posttraining, while the maximal group mean decreased from training to posttraining.

Overall, the means were higher later in the year (contributing to the main effect for time of year). In general, these two variables show that some differential use of large group instruction was occurring across the treatment groups according to time of year. This may be noteworthy, since the other analyses--analysis of variance and analysis of covariance--did not suggest any treatment effects on measures of large group instruction.

The last variable in Table 8 is the measure of adult movement. This variable shows a treatment main effect and a time-of-year main effect, but no treatment by time-of-year interaction. The direction of the means indicates that only the minimal and control groups were able to maintain or increase their baseline rates during training and posttraining periods. The maximal training group decreased from baseline to training and then increased to posttraining. However, these differences in patterns between groups were not large enough to cause a treatment by time-of-year interaction. The time-of-year main effect is primarily caused by the increase from training to posttraining that occurred consistently across all treatment groups.

The repeated measures ANOVAs provide a great deal of useful information. About half of the 23 FMO variables showed treatment main effects and slightly less than 40% yielded evidence of treatment by time-of-year interactions. Although not all of the interactions were interpretable and consistent with the training recommendations, most of the 40% showing significant or near-significant interactions did present evidence that the experimental groups' behavior was more in accord with the training recommendations after baseline than was the behavior of the control group. Analysis of a few variables indicated that a treatment main effect on training and posttraining phase data was reflected in similarly oriented differences during baseline. While it is possible that the experimental intervention maintained

or enhanced those initial differences when they might have otherwise disappeared, in general we should not claim strong treatment effects for variables showing baseline differences in the same direction as post-baseline data. The use of random assignment, particularly with a small number of subjects, will not preclude initial group differences on all measures of interest.

Assessment of Process Variation related to the Students' Characteristic "Academic Orientation"

As was briefly noted earlier, some of the Questioning and Feedback category recommendations were worded in such a way that teachers were advised to teach differently to different types of students. Behavioral descriptions of what was meant by more or less "academically-oriented" students were given to experimental group teachers in the first training packet, and they were then asked to rate each of their students on a 4-point scale from "Extremely Academically Oriented" to "Extremely Non-Academically Oriented." For the purpose of the analyses reported here, the top two categories ("extremely" and "somewhat" academically oriented) were combined, and the bottom two categories ("extremely" and "somewhat" non-academically oriented) were combined. This yielded a dichotomous "more" vs. "less" academically-oriented distinction. Table 9 below shows the breakdown of number of more vs. less academically-oriented students in each of the 23 experimental group (maximal and minimal training) teachers' classes.

Insert Table 9 about here

From this table it can be seen that most of the teachers rated more than half of their students as more academically-oriented. The overall mean for the variable "Percentage of students rated as more academically oriented" was

67.8%, with a mean for District I of 75.8%, and a mean for District II of 59.1%. It is possible that the teachers' ratings are somewhat higher than they should be, but there is a large degree of accuracy in these ratings as well. The analysis of statewide norms reported in PTE (1978) showed the two sample districts to be higher than the overall state population on measures of SES. The behavioral definitions of academic orientation were based on SES differences on process variables (as presented in Brophy, Evertson, Crawford, & Sherman, 1975). The agreement of teachers' percentages with statewide data seems a noteworthy validation of the teachers' ratings of behaviorally-defined student characteristics.

The most important question about implementation of the SES-differentiated behavior is whether or not the teachers receiving training were able to implement effectively the different teaching methods with different kinds of students. The method of analysis used to examine this issue was to correlate the variable "% of students rated more academically-oriented" with relevant process measures within the sample of experimental group classes (N = 23; control group classes did not rate their students' academic orientation).[2] In this way, we would be able to determine whether rates of certain process variables covaried significantly with the academic orientation makeup of the classes.

In general, the different recommendations for teaching different types of students call for methods likely to maintain a high degree of student motivation and interaction with the academic material to be learned. For less

---

[2]It had been originally planned to use the low inference child focus data to address these questions--and to conduct analyses with the student, not the class, as the unit of analysis. However, because of the large number of students in the sample and because of the higher priority assigned to obtaining teacher focus data, only 4 FMOs per child were possible. The distributions (and low reliabilities) of the variables related to teaching more vs. less academically oriented students were such that the use of these measures for inferential analyses was not considered possible. (The distributions were highly skewed; several variables had a mode of 0.0, with 1% or less of the students determining the variance.) Hence, a decision was made to use the class as the unit of analysis in conjunction with the "% more academically oriented" index.

academically-oriented students, teachers were asked (1) to use sustaining (more lengthy) feedback when necessary, (2) to avoid allowing students to passively respond "I don't know" or make no response at all, (3) to engage these children in some personal, non-academically related discussions in order to maintain rapport and promote student involvement, (4) to give praise when warranted and avoid the use of criticism, (5) to refrain from discouraging student "call outs," and (6) to ask questions that can be answered with a high rate of success. When teaching more academically-oriented students, teachers were advised (1) to keep discussions fast-paced through the use of brief feedback, (2) to focus on academic interactions and not encourage lengthy personal discussions with students, (3) to give praise very sparingly and to make use of some criticism to communicate high expectations, (4) to tell students not to call out answers, but to wait until called on or until their turn (the "student call outs" variable had negative correlations with achievement for the total sample of classes in the Brophy-Evertson study, but the rs for the high SES subsample were consistently even more negative than for lower SES classes), and (5) to increase the difficulty of questions so that students are challenged and make some mistakes (an error rate of about 25% was recommended).

The correlations between relevant process measures and the percentage of students rated as more academically-oriented are presented in Table 10 below. High inference variables were rated after each morning and each afternoon of observation during the posttraining period, so separate correlations are given for morning and afternoon data. For low inference variables, correlations are presented separately for the training and posttraining phases. Since the districts differed in the overall percentages of more and less academically-oriented

students, correlations were also computed within districts, in addition to being presented for the total sample of experimental group classes (N = 23).

Insert Table 10 about here

If teachers in classes with lower percentages of more academically-oriented students (i.e., with relatively large numbers of less academically-oriented students) followed the training recommendations about using personal discussions and non-academic questions to maintain student involvement, then we should find negative correlations in Table 10 for the high inference measures of "Personal, private interactions," and for the low inference variables reflecting "self-related" questions ("self-related" means questions from the teacher about the child's personal, often not school-related, concerns--e.g., inquiries about home projects or parents). The results in Table 10 support this hypothesis, and also show that the effect is consistently stronger in District II, the district with the higher percentage of less academically oriented students.

The training materials and workshops also emphasized the need for promoting responses from less academically-oriented students. For example, these students should not be allowed to remain silent or say only "I don't know" when asked a question. The recommendation implies, then, that successful implementation would yield classes with higher percentages of less academically-oriented students having relatively low ratings on "pupil passivity." That is, the correlations for "% more academically-oriented" and passivity in Table 10 should be positive. The findings are mixed for this variable. For passivity

in the morning, the total sample correlation is slightly negative ($r = -.06$), but the within-district correlations are slightly positive for both districts ($rs = .11$ and .15 for Districts I and II, respectively). The fact that two positive within-district rs can combine to give an overall negative correlation reflects a degree of non-overlap between the ranges of the within-district distributions on one or both of the variables entering into the correlation. The afternoon ratings of passivity also show a negative correlation for the total sample, and a positive r for District I and a negative r for District II. The findings present only weak support for the hypothesized relationships concerning the measures of passivity, and even then, only in the within-district analyses.

Recommendations concerning the differential use of praise and criticism may be assessed through analysis of ratings of positive affect (as a proxy for praise) and negative affect (a proxy for criticism). If teachers used more praise with less academically oriented students and more criticism with more academically-oriented students, then one would expect the correlations of positive affect with "percentage more academically-oriented" to be <u>negative</u> and <u>positive</u> for negative affect. The rs for positive affect are consistently in the right direction (negative) for both morning and afternoon data. The relationships are most pronounced for the District I subsample ($rs = -.35$ and $-.41$ for A.M. and P.M. data, respectively). However, the negative affect measure did not show the expected pattern of (positive) correlations. The total sample of rs were negative in both the morning ($r = -.11$) and afternoon ($r = -.23$) data. The within-district correlations in the morning were slightly positive ($rs = .05$ and .06), but should not be considered meaningfully different from zero.

The remaining two variables in Table 10 to be discussed are low inference measures of student "call outs" during public questioning sessions and the teachers' use of rephrasing or clueing after incorrect responses. In the prior research (Brophy and Evertson, 1974), "call outs" was strongly negatively correlated with achievement in the high SES group of classes. Therefore, the recommendation to teachers in this experiment was that they should actively discourage "call outs" from more academically-oriented students. If they adhered to this recommendation, then the rs in Table 10 for "call outs" should be negative. For this variable measured during the training phase, a slight negative r of -.04 for the total sample was found, along with a stronger negative r of -.18 for District I and a weak positive r of .07 for District II. The post-training measure also showed a weak negative total sample r of -.07, but this was associated with two positive, though slight, correlations when computed separately within districts.

The "rephrases or gives clue" variable should have been negatively related to "percentage of students rated more academically-oriented" if the trained groups were directing more of this kind of feedback to the less academically-oriented students. However, the correlations are consistently positive for this measure in both the training and posttraining phases, and both for the total sample, and the separate districts. This is the only one of the implementation variables in Table 10 that yielded no evidence of successful differentiation of teaching according to the students' academic orientation. It may be that the rapid teacher decision-making requirements during questioning sessions do not permit selective use of rephrasing and clueing strategies. Or, it is possible that there is a problem in measuring this Brophy-Evertson variable with the modified COI system. It may even be that a number of teachers in primarily "more academically-oriented"

classes were effectively using rephrasing and clueing with a minority of the students in their classrooms. If this tendency occurred consistently and often enough, the correlations based on class means could be opposite in sign from what was expected.

These correlational analyses indicate some degree of teacher success in differentiating teaching style according to behaviorally-defined student characteristics. The findings are not unequivocal, and this is not the only method of analysis that could have been used in addressing this issue. However, even such tentative confirmation of this aspect of the teacher training program is notable, since the "academically-oriented" based recommendations appear to go a significant step beyond the majority of "individualization" programs--those that typically rely on teacher use of small group and one-to-one instruction, with little or no advice on how to instruct. This kind of experimentation seems required if the suggestions of aptitude-treatment interactions, as in the Brophy-Evertson results, are ever to be integrated into pre- or inservice teacher education programs.

Confirmation of Prior Process-Product Correlational Results

The most interpretable and consistent findings of the four previous correlational studies were incorporated into the teacher training program. However, given differences in samples of teachers and students, there was no guarantee that all of the selected behaviors would relate to achievement in this study in the same direction and to the same degree as in the previous studies. A major question of interest, then, concerns the consistency with which process-product correlations from the present sample agree with those from the earlier studies.

The synthesis of findings across different studies has only recently received serious attention from social and behavioral scientists. Glass (1977) and Gage

(1976) have discussed techniques for combining results from different investigations. We have derived specific criteria for considering a prior correlation coefficient to be confirmed. The major criterion in this analysis is that two of the three correlations with achievement outcomes (adjusted vocabulary, adjusted reading comprehension, and adjusted total reading) in this study must agree in sign with the previous study's process-product r, and at least one of those two rs must be equal to or greater than .15 in magnitude. A few variables (three) were also considered to show some evidence of confirmation of prior results because all three correlations with achievement agreed in sign with the previous study's coefficient, even though none of the three in this study was equal to or greater than .15.

In deriving a "cutoff" magnitude of .15, our reasoning was as follows:

> The correlations from the previous studies were associated with p-values of .05 to .10 or less. In this study, with $N = 33$, an r of .15 has a probability of about .20. Applying the logarithmic transformation described by Jones and Fiske (1953) to independent p-values of .20 (this study) and .10 (previous study) yields:
>
> $$\chi^2_{(\text{combined})} = -2\sum_{i=1}^{k} \log_e p_i, \text{ where}$$
>
> k = number of p-values (i.e., number of studies) = 2
>
> p_i = p-values from each of the independent studies; in this illustration, $p_1 = .20$ and $p_2 = .10$.
>
> Therefore, $\chi^2_{(\text{combined})} = 7.824$ with $2k$ (=4) df
>
> and: $\chi^2 = 7.824$ and df = 4, $.05 < p \leq .10$.

So, by using a criterion of $r = .15$ in this study as evidence of confirmation, we are running only a 5 to 10% risk of claiming agreement between studies where none exists. This derivation of a single criterion of $r = .15$ for all

variables is actually somewhat conservative, since most of the rs selected from previous studies had p-values less than .05 (if p_2 in the above computation were .05, then the $\chi^2_{(combined)}$ would be larger).

It should also be noted that all of the correlations being considered have been attenuated because of imperfect reliability of the process variable and the achievement measure entering into the correlation. All coefficients would be increased in magnitude if corrections for this attenuation were employed. For example, suppose a typical process variable has a reliability of .70. (This was the average of the reliabilities computed for low inference process variables in this study; see Appendix D in PTE, 1978.) Assume a reliability of adjusted achievement of .60.[3] If we then use the "correction" formula from Spearman:

$$r_{disatten.} = \frac{r_{observed}}{\sqrt{Xreliability * Yreliability}}$$

and take a typical process-product r of .20, we find that,

$$r_{disatten.} = \frac{.20}{\sqrt{(.70) * (.60)}} = .31$$

In this example, then, the r would be increased from .20 ($p > .10$) to .31 ($p < .05$) by the disattenuation procedure. Available time and resources

---

[3] In this study the reliability of the unadjusted class means is >.90 for the achievement variables. However, the regression procedure that removes pretest and district variation will result in lower reliabilities for the adjusted scores, since the remaining variation will consist of a larger relative proportion of variance due to error (larger than the proportion of variance due to error in the unadjusted class means). We have not yet computed the reliability of adjusted outcomes and offer the figure of .60 as purely speculative, for illustrative purposes, to show the effects of disattenuation.

do not permit a disattenuation study of all the process-product correlations, but we provide this apparently typical example as evidence of how much the "corrected" relationships might increase. The example might be informative for those readers inclined to infer, from the small number of statistically significant process-product rs, that teaching behavior correlates of student achievement are negligible. The major point to be made from the discussion of combining results through χ^2 analysis and the example of a disattenuated correlation is that we are more likely to be making Type II than Type I errors when drawing inferences from the process-product matrix presented in Table 11. (The application of Cohen's, 1977, power formulas yields the same conclusion.)

Insert Table 11 about here

In view of the problems of measurement and statistical power, the results presented in Table 11 are about as encouraging as one might expect. Further confirmation of the present sample's high SES is given by the agreement between the process-product rs in this study and the Brophy-Evertson high SES rs. Overall, about half of the variables in Table 11 show some evidence of agreement with the previous correlational results. By training category, the data suggested that (1) 6 of 11 previous findings are confirmed in the Behavior Management area, (2) 9 of 24 previous results are confirmed in the Instructional Methods area, and (3) 8 of 11 previous correlations are confirmed in the Questioning and Feedback area, when considering this sample to be comparable to the Brophy-Evertson high SES group. The low proportion of relationships confirmed in the Instructional Methods category may have been caused by the fact that most of those variables come from the study by Stallings and Kaskowitz which was conducted in primarily lower SES (Follow Through) schools. The Brophy-Evertson findings and our own results

indicate that the SES level of the students can have substantial influence on the effectiveness of certain teaching strategies. The results for the one-half of the variables on which our findings did agree with the previous findings are important to the degree that they demonstrate that certain components of teaching behavior relate to student achievement in ways that generalize across independently conducted investigations.

Conclusions

The analysis of observational data presented in this paper has provided information useful in the evaluation of the teacher training program and has also pointed to a number of analysis and measurement issues that seem particularly relevant to the study of teaching in public school classrooms. One conclusion relates to the adequate measurement of process variables from different coding systems with the very general time-sampling instrument used in this study: the SRI Classroom Observation Instrument (COI). Our findings suggest that we should have developed specific coding procedures and categories for assessing implementation of treatment variables originating in studies other than that of Stallings and Kaskowitz (1974) (which also used the COI). This approach would have served better than relying solely on post hoc creation of sequences of codes that only nominally represent the behavior of interest. As long as coders remain "blind" to the treatment group assignment of individual teachers, the implementation data should not be biased by the training of observers to record data for specific treatment-related observation variables. Our study suggests that even a highly developed, well-documented low-inference system like the COI in the hands of trained observers cannot always be depended upon to yield accurate measurement of variables originating in other systems. The high "percentage agreement" figures in Table 2 indicate that the coders were well-trained in using the individual categories of the system.

We have chosen α-levels higher than the usual .01 and .05 as suggestive of effects worth some discussion and interpretation. This experiment is a first step in the direction of relatively rigorous study of causal influences of training on classroom behavior and of classroom processes on student learning. Random assignment was used, and the teacher training treatment has direct ties

to earlier empirical work. But since it is a first step, Type II errors deserve as much consideration as Type I errors in our judgment. Even allowing p-levels of .10 to .20 to be considered as near significant, we are still probably missing many relationships that "exist" (i.e., that would be found if the experiment were repeated a large number of times, or if the N of classes in the study were increased substantially). Unfortunately, this type of research is expensive, even in a sample of 33 classes. Hence, it seems unlikely that resources will permit such intensive experimentation in classes numbering 100 or more per treatment group.

Although ANOVA and ANCOVA results are reported, we have attempted to demonstrate that the assessment of implementation of naturally occurring treatments is not always a between-group question. Still, 7 of 16 components were implemented correctly and showed significant or near-significant between-group differences (and 10 of 16 component means differed in the correct direction--with control group classes having the lowest implementation score). The treatment by time-of-year interaction source in the repeated measures analyses of variance seems to us, in many ways, to be the "most" appropriate indication of treatment effects in an experiment of this type. Accordingly, we would strongly recommend that researchers engaged in this kind of study obtain measures of implementation variables before, during, and after the treatment intervention so that treatment-by-phase effects can be studied (rather than measuring implementation only after the treatment and depending on random assignment or matching as insurance against initial group differences).

The repeated measures ANOVAs also showed that more variables evidenced time-of-year main effects than treatment main effects or interactions. This

result indicates that, while the intervention did causally influence many classroom processes in the expected fashion, other forces affecting treatment behaviors were operating during the experiment. Since the study was conducted in public school classes with regular teachers, it was not possible to control all of these factors. We did, however, design measurement procedures to allow estimation of the magnitude of these effects and thereby afford the opportunity of documenting that which we could not control. Even with the possible confounding of these uncontrolled factors, this kind of research seems required to determine the effects of inservice teacher education of the kind employed in this experiment. Factors such as time-of-year and variation in teacher attributes would probably affect the implementation of district or state-mandated teacher training programs--as was the case in the PTE study with our experimental program.

The evidence indicates that trained teachers did not follow all recommendations to the same degree. It was also true, though, that the behaviors did not always relate to student achievement in the same direction or to the same degree in this study as in the previous research. The training-group teachers may have focused selectively on implementation of behaviors that they believed would enhance student learning in their students. It would be possible to avoid training teachers in potentially ineffective strategies by establishing the process-product basis of the training in the <u>same sample</u> of classes and teachers that is to receive the training. Although such a study lacks the external validity of an experiment that draws on a wide range of sampled classes and behaviors as its empirical base, the treatments derived in this fashion would be more "powerful" in the sense of being known with a greater degree of certainty to relate to

enhanced achievement in the sample of interest. Jane Stallings of SRI International is presently conducting a study of this type.

The present data do not provide confirmation of all expectations. Yet, the findings seem promising enough to confirm the utility of this "empirically-based-intervention" paradigm. Future studies providing experimental tests of correlational findings should be able to use measurement and training techniques improved in the light of the present results.

References

Brophy, J., & Evertson, C. Process-product correlations in the Texas Teacher Effectiveness Study: Final Report. Research Report No. 74-4. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1974.

Brophy, J., Evertson, C., Crawford, J., & Sherman, G. The Texas Teacher Effectiveness Study: Student sex, grade, and socioeconomic status differences in classroom process measures. Report No. 75-21. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1975.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences, (Rev. ed.). New York: Academic Press, 1977.

Corno, L. Effects of Parent Instruction on Teacher Structuring and Student Participation in the Third Grade: An Aptitude-Treatment Interaction Approach. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada, March 1978.

Crawford, J., & Stallings, J. Experimental Effects of In-Service Teacher Training Derived from Process-Product Correlations in the Primary Grades. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada, March 1978.

Gage, N. L. Four Cheers for Research on Teaching. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, 1976.

Glass, Gene V . Integrating findings: The meta-analysis of research. Review of Research in Education, 1977 (in press).

Jones, L. V., & Fiske, D. W. Models for testing the significance of combined results. Psychological Bulletin, 1953, 50, 375-382.

Kerlinger, F.N., & Pedhazur, E.J. Multiple regression in behavioral research. New York: Holt, Rinehart and Winston, 1973.

Kounin, J. Discipline and group management in classrooms. New York, N.Y.: Holt, Rinehart & Winston, 1970.

McDonald, R., & Elias, P. The effects of teaching performance on pupil learning, Beginning Teacher Evaluation Study: Phase II, Final Report: Vol. I. Princeton, N.J.: Educational Testing Service, 1976.

Program on Teaching Effectiveness. An experiment on teacher effectiveness and parent-assisted instruction in the third grade: Volume I. Center for Educational Research at Stanford, 1977. National Institute of Education Contract No. NE-C-00-3-0061.

Program on Teaching Effectiveness. An experiment on teacher effectiveness and parent-assisted instruction in the third grade: Volume III. Center for Educational Research at Stanford, 1978. National Institute of Education Grant No. OB-NIE-G-78-0112.

Salomon, G. Heuristic models for the generation of aptitude-treatment interaction hypotheses. Review of Educational Research, 1972, 42, 327-343.

Soar, F. Final report, Follow Through classroom process measurement and pupil growth (1970-1971). Gainesville, Fla.: Institute for the Development of Human Resources, University of Florida, 1973.

Stallings, J., & Kaskowitz, D. Follow Through classroom observation evaluation, 1972-1973. Menlo Park, Calif.: Stanford Research Institute, 1974. Office of Education Contract OEC-0-8522480-4633(100).

Ward, J.H., & Jennings, E. Introduction to linear models. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

TABLE 1

Definitions, Sources, and Expected Direction of Process-Product Correlation for Process Variables

Variable #	Variable Name[1]	Instrument for measurement in this study.[2]	Direction of r in prior study.
	Behavior Management Process Variables		
BM01	Use of a system of rules to minimize interruptions (BE).	HI	+
BM02	Timing errors--allowing misbehaviors to continue too long (BE).	HI	-
BM03	Overreactions--teacher use of stronger-than-necessary discipline (BE).	HI	-
BM04	Adult attention to large group (SK).	FMO	+
BM05	Adult movement (SK).	FMO	+
BM06	Checking (teacher activities)(McE).	FMO	+
BM07	Supervising (teacher activities)(McE).	FMO	-
BM08[3]	Target errors--directing discipline at the wrong child (BE).	HI	-
BM09[3]	Awareness of behavior problems.	HI	+
BM10[3]	Communication of awareness to students (together with BM09 constitutes a measure of teacher "withitness").	HI	+
BM11[3]	Direction giving.	HI	-
	Instructional Methods Process Variables		
IM01	Free time materials available (not necessarily used): games (BE).	CS	-
IM02	Ratio of self-related (personal) questions to academic questions (BE).	FMO	-
IM03	Ratio of opinion questions to academic questions (BE).	FMO	-
IM04	Games, toys, play equipment present (SK).	CS	-
IM05	Games, toys, play equipment used (SK).	CS	-

TABLE 1 (continued)

Definitions, Sources, and Expected Direction of Process-Product Correlation for Process Variables

Variable #	Variable Name[1]	Instrument for measurement in this study.[2]	Direction of r in prior study.
	Instructional Methods Process Variables (cont'd)		
IM06	Blocks, trucks, activities (SK).	CS	-
IM07	Teacher with small groups (SK).	CCL	-
IM08	Adult with large group (SK).	CCL	+
IM09	Large group of children independent in reading (SK).	CCL	+
IM11	Adult commands, requests, and directs questions to individual children, non-academic (SK).	FMO	-
IM12	Adult commands, requests, and directs questions to groups of children, academic (SK).	FMO	+
IM13	Adult academic commands, requests, and directs questions to children (SK).	FMO	+
IM14	Child response, non-academic (SK).	FMO	-
IM15	Adult instruction, academic (SK).	FMO	+
IM18	Individual child verbal interaction with adult (SK).	FMO	-
IM19	Adult communication or attention focus, large group (SK).	FMO	+
IM20	Phonics skills practice (pupil events) (McE).	CS	+
IM21	Academic organizing (teacher activities) (McE).	CCL	-
IM23	Instructing (teacher responses)(McE).	FMO	+
IM24	Materials: board and related equipment (McE).	CS	+
IM25	"Secondary" materials: books (McE).	CS	+

TABLE 1 (continued)

Definitions, Sources, and Expected Direction of Process-Product Correlation for Process Variables

Variable #	Variable Name[1]	Instrument for measurement in this study.[2]	Direction of r in prior study.
	Instructional Methods Process Variables (cont'd)		
IM26	Activity: discussion (McE).	CCL	+
IM27	Activity: seatwork (McE).	CCL	-
IM28	Activity: audiovisual aids (McE).	CCL	+
IM29	Activity: visual demonstration (McE).	CS	+
IM30	Teacher structured activity (vs. pupil free choice) (S).	CS	+
IM31	Taxonomy of cognitive behavior: information giving and receiving (S).	CCL	-
	Questioning and Feedback Process Variables		
QF01	Percent of preselects or calls on non-volunteers before asking question (BE).	FMO	+
QF02	Percent of student call-outs in response to questions (BE).	FMO	-
QF03	Quality of child's answer: Percent correct (BE).	FMO	+ for Lo SES - for Hi SES
QF04	Quality of child's answer: Percent wrong (BE).	FMO	- for Lo SES + for Hi SES
QF05	Quality of child's answer: Percent no response (BE).	FMO	- for Lo SES
QF07	Teacher gives the answer after part-correct response from child (BE).	HI	+
QF08	Teacher criticism after wrong answer (BE).	FMO	- for Lo SES + for Hi SES
QF09	Teacher rephrases or gives clues after wrong answer (BE).	FMO	+ for Lo SES - for Hi SES

TABLE 1 (continued)

Definitions, Sources, and Expected Direction of Process-Product Correlation for Process Variables

Variable #	Variable Name[1]	Instrument for measurement in this study.[2]	Direction of r in prior study.
	Questioning and Feedback Process Variables (cont'd)		
QF10	Teacher sustaining feedback after child responds "I don't know" or makes no response (BE).	FMO	+ for Lo SES - for Hi SES
QF11	Reactions combined across all response opportunities: teacher criticism (BE).	FMO	- for Lo SES + for Hi SES
QFI3	Percent student opinions accepted (BE).	FMO	-
QF19	Teacher feedback combined across response opportunities: rephrase ÷ repeat + new question (BE).	FMO	-
QF21	Child action directed to adult: all verbal actions except responses (SK).	FMO	-
	Low Inference FMO Variables with Insufficient $\hat{\sigma}^2$(class) for Analyses of Effects		
IM10	Child commands, requests, and directs questions, nonacademic (SK).	FMO	-
IM16	Child attentive to machine (SK).	FMO	-
IM17	Self-instruction, academic (SK).	FMO	+
IM22	Instruction giving (teacher activities) (McE).	FMO	-
QF06	Process (detailed) feedback after student correct answer (BE).	FMO	+
QF12	Process (detailed) feedback after all student responses (BE).	FMO	+
QF14	Private dyadic contacts: Percent student initiated contacts given brief feedback (BE).	FMO	- for Lo SES + for Hi SES

TABLE 1 (continued)

Definitions, Sources, and Expected Direction of Process-Product Correlation for Process Variables

Variable #	Variable Name[1]	Instrument for measurement in this study.[2]	Direction of r in prior study.
	Low Inference FMO Variables with Insufficient $\hat{\sigma}^2$ (class) for Analyses of Effects (cont'd)		
QF15	Private dyadic contacts: teacher initiated work contacts involving praise (BE).	FMO	+ for Lo SES - for Hi SES
QF16	Private dyadic contacts: teacher initiated work contacts involving mere observation (BE).	FMO	-
QF17	Private dyadic contacts: teacher initiated work contacts involving long feedback (BE).	FMO	+ for Lo SES - for Hi SES
QF18	Private dyadic contacts: student initiated work contacts involving praise (BE).	FMO	+ for Lo SES - for Hi SES
QF20	Combined teacher feedback: brief ÷ brief + long feedback (BE).	FMO	- for Lo SES + for Hi SES
QF22	Child general comment to adult (SK).	FMO	-

[1]At the end of each variable name the study in which the variable originated is indicated: BE--Brophy and Evertson, SK--Stallings and Kaskowitz, McE--McDonald and Elias, and S--Soar.

[2]FMO: Five Minute Observation from the low inference COI.
CCL: Classroom Checklist from the low inference COI.
HI: High Inference ratings.
CS: Curriculum Survey.

[3]These variables were included in the training program as integrative concepts, and not because there were identifiable process-product correlations in the earlier studies.

Table 2

Average Interrater Agreement Percentages for the Five-Minute Observation Codes

Five-Minute Observation Code	Average Percent Interrater Agreement
WHO:	
T - Teacher	82.34
A - Adult	100.00
V - Volunteer (child)	99.64
C - Child	87.69
D - Different child	98.49
2 - Two Children	100.00
S - Small Group (3-8)	98.91
L - Large Group	98.58
An - Entire Class	99.06
M - Machine	100.00
TO WHOM:	
T - Teacher	88.43
A - Adult	99.79
V - Volunteer (child)	99.79
C - Child	79.97
D - Different Child	96.67
2 - Two Children	99.95
S - Small Group (3-8)	98.09
L - Large Group	92.35
An - Entire Class	90.58
M - Machine	98.58
WHAT:	
1 - Command or Request	92.25
2 - Open-ended Question	99.90
3 - Response	93.58
4 - Instruction, Explanation	93.52
5 - General Comments or Action	92.96
6 - Task-related Comments	91.18
7 - Acknowledgement, Encouragement	96.51
8 - Praise	99.48
9 - Corrective Feedback	96.32
10 - Process Feedback	99.95
11 - No Response, Don't Know	98.34
12 - Observing, Listening, Monitoring	90.03
HOW:	
NV - Nonverbal	84.63
X - Movement	95.75
H - Happy	93.81
U - Transition Activities	94.90
N - Negative	95.54
T - Touch	99.85
Q - Question	93.01
G - Guidance	99.22
P - Incorrect Response	98.39
O - Organizing, Structuring	99.78
W - Statements of Self-Worth	99.79
Dp - Length of Interaction	94.48
A - Academic	72.32
B - Behavior	99.59

Table 3

Average Interrater Agreement Percentages
for the High Inference Variables
(Based on 3 Days of Observation)

Variable Number[1]	Average Interrater Agreement
BM01	92%
BM02	83%
BM03	100%
BM08	92%
BM09	75%
BM10	75%
BM11	100%
QF07	75%

[1]See Table 1 for definitions of variables.

TABLE 4

Specific Variables forming Composites that represent Components of the Training Program

Variable #	Name	Specific Variable entering Composites, Weights indicated
Behavior Management:		
BM1	Use of a system of rules.	BM01(+)[1]
BM2	Timing errors.	BM02(-)
BM3	Overreactions.	BM03(-)
BM4	Target errors.	BM08(-)
COMPBM5[2]	"Withitness".	BM09(+), BM10(+)
COMPBM6	Monitoring.	BM05(+), BM06(+), BM07(-)
General Instructional Methods:		
COMPIM1	Children independent, academic.	IM09(+), IM27(-)
COMPIM2	Direction giving, organizing.	IM21(-), IM31(-), BM11(-)
COMPIM3	Use of large groups.	IM08(+), IM19(-), BM04(+)
COMPIM4	Phonics, visual and audiovisual aids.	IM28(+), IM28(+), IM29(+)
COMPIM5	Academic teacher-student interactions.	IM13(+), IM12(+), IM23(+), IM26(+)
COMPIM6	Non-academic teacher-student interactions.	IM11(-), IM02(-), IM14(-), IM18(-), QF21(-)
Questioning and Feedback:		
QF1	Preselecting or calling on nonvolunteers.	QF01(+)
QF2	Student "call outs".	QF02(-)
QF3	Percentage of answers wrong.	QF04(+)
QF4	Teacher gives right answer after student part-correct response.	QF07(+)

[1]See Table 1 for definitions of specific variables. The sign in parentheses indicates the direction of successful implementation. The weights (+1.0 or -1.0) were multiplied by the scores of the variables, and then Z-scores were created and summed for composites.

[2]"COMP" stands for composite, indicating more than 1 variable entering into the implementation score.

TABLE 5

Summary of Analysis of Variance of Components of the Teacher Training Program

Variable	Main Effect F-ratio	p	Direction of Means (Scores have been oriented so that a high score means high implementation)
*BM1--Use of a system of rules.	2.590	.094	Minimal > Maximal > Control
*BM2--Timing errors.	2.399	.110	Minimal > Maximal > Control
BM3--Overreactions.	0.273	.763	Maximal > Control > Minimal
BM4--Target errors.	1.110	.344	Minimal > Control > Maximal
*COMPBM5--"Withitness".	2.832	.076	Minimal > Maximal > Control
COMPBM6--Monitoring.	2.199	.130	Control > Maximal > Minimal
COMPIM1--Children independent, academic.	0.787	.465	Minimal > Maximal > Control
*COMPIM2--Direction giving, organizing.	4.605	.019	Minimal > Maximal > Control
COMPIM3--Use of large groups.	0.452	.641	Minimal > Maximal > Control
COMPIM4--Phonics, visual and audiovisual aids.	1.124	.340	Minimal > Maximal > Control
COMPIM5--Academic teacher-student interactions.	0.274	.763	Control > Minimal > Maximal
COMPIM6--Non-academic teacher-student interactions.	0.633	.539	Minimal > Control > Maximal
*QF1--Preselecting or calling on nonvolunteers.	4.566	.020	Maximal > Minimal > Control
*QF2--Student "call outs".	3.982	.031	Maximal > Minimal > Control
*QF3--Percentage of answers wrong.	1.948	.162	Minimal > Maximal > Control
QF4--Teacher gives right answer after student part-correct response.	0.836	.444	Minimal > Control > Maximal

*Indicates successful implementation considering 2 criteria: (1) a significant or near-significant F-ratio (with α-level at .10), and (2) Control group mean is the lowest with either the minimal or the maximal training group having the highest mean.

TABLE 6

Regression Statistics for Homogeneity of Regression of Low Inference FMO and Classroom Checklist Variables

Variable #[1]	r^2_{xy}[2]	District I Slopes			District II Slopes			Pooled Slope	F[3]
		Maximal	Minimal	Control	Maximal	Minimal	Control		
QFO1	.78	.18	.75	.73	.49	.82	.79	.70	0.22
QF02	.45	.55	.32	1.22	-.04	.52	.76	.44	2.02*
IMO2	.40	.68	.08	-.49	.23	.65	.23	.31	1.09
BMO6	.71	1.22	.54	-.41	.97	.69	1.04	.69	1.82*
BMO7	.51	.51	1.09	-.93	1.70	1.13	.89	.86	1.87*
IM23	.54	.55	.86	.18	.70	-.75	.98	.52	1.86*
IM11	.53	.66	.39	.12	.19	.43	.03	.40	0.01
IM12	.56	.40	.35	1.51	.40	1.04	-.48	.50	1.53
IM13	.79	.64	.79	.99	.46	.67	.72	.69	0.01
IM14	.48	.71	.76	.04	.24	.28	.47	.48	0.72
IM18	.63	.25	.50	-.17	1.09	.33	.77	.41	1.25
QF21	.77	.84	.59	.43	.76	1.65	1.29	.79	0.48
IM19	.43	.30	.69	.57	.82	.08	1.23	.57	0.55
BMO4	.50	.61	.61	.21	1.11	.19	.74	.54	0.34
BMO5	.58	.44	.71	.52	.34	.69	1.29	.51	0.41
IMO7	.37	-.19	-.19	.97	1.04	.81	.83	.50	2.14**
IMO8	.45	.63	.97	-.10	-.18	.77	.62	.46	0.75
IM21	.37	.25	.33	.05	.15	1.21	.47	.25	0.54
IM26	.62	.38	.72	-.25	.27	.61	.52	.48	0.95
IM27	.56	.51	.57	.38	.21	.44	.71	.43	0.23
IM31	.46	-.25	-.32	.32	.27	-.10	.34	.24	0.46

* $.10 < p \leq .20$

** $p \leq .10$

[1] See Table 1 for definitions.

[2] r_{xy} = r between baseline and post-baseline (training and posttraining aggregated).

[3] A significant F-ratio indicates heterogeneity of regression coefficients across groups (ATI).

TABLE 7

Adjusted and Raw Means for Discrete Process Variables

Variable	Raw Means			Means Adjusted for Baseline Rates			F ratios for Treatment main effect
	Max	Min	Cont	Max	Min	Cont	
BM04--Adult attention to large group (SK).	.0346	.0358	.0346	.0346	.0391	.0320	.716
BM05--Adult movement (SK).	.0746	.0687	.0816	.0701	.0753	.0798	1.314
IM02--Ratio of self-related (personal) questions to academic questions (BE).	.1413	.1139	.1303	.1310	.1132	.1471	.545
IM08--Adult with large group (SK).	.7450	.7491	.6879	.7666	.7218	.6919	.592
IM11--Adult commands, requests, and directs questions to individual children, non-academic (SK).	.0098	.0084	.0096	.0093	.0081	.0107	.820
IM12--Adult commands, requests, and directs questions to groups of children, academic (SK).	.0123	.0134	.0168	.0120	.0129	.0174	1.534
IM13--Adult academic commands, requests, and directs question to children (SK).	.0483	.0523	.0437	.0459	.0473	.0509	.977
IM14--Child response, non-academic (SK).	.0447	.0446	.0529	.0447	.0443	.0537	1.749*
IM18--Individual child verbal interaction with adult (SK).	.1439	.1373	.1252	.1372	.1344	.1364	.154
IM19--Adult communication or attention focus, large group (SK).	.1858	.2022	.1794	.1889	.2062	.1688	1.799*

TABLE 7 (continued)

Adjusted and Raw Means for Discrete Process Variables

Variable	Raw Means			Means Adjusted for Baseline Rates			F ratios for Treatment main effect
	Max	Min	Cont	Max	Min	Cont	
IM21--Academic organizing (teacher activities) (McE).	1.4824	1.0660	1.7266	1.4929	1.1655	1.6046	1.016
IM26--Activity: discussion (McE).	1.4324	1.3782	1.8430	1.3882	1.3908	1.8821	1.166
IM27--Activity: seatwork (McE).	8.9843	9.2186	10.6598	8.8747	9.5352	10.4432	1.482
IM31--Taxonomy of cognitive behavior: information giving and receiving (S).	.8959	.5492	1.0404	.9110	.6542	.9068	1.387
QF01--Percent of preselects or calls on nonvolunteers before asking question (BE).	.6296	.6282	.5123	.5929	.6050	.5769	.448
QF21--Child action directed to adult: all verbal actions except responses (SK).	.0635	.0516	.0590	.0575	.0572	.0606	.224

* $.10 < p \leq .20$

TABLE 8

Repeated Measures Means and F-Ratios for Discrete Process Variables: Baseline, Training, and Posttraining Phases

Variable	F-ratios F_{Trmt}	$F_{Time\ of\ Yr}$	$F_{Trmt \times Time}$	Means Time of Year	Maximal	Minimal	Control
Selection of respondents to questions: proportion preselects respondents or calls on non-volunteers before asking questions.	4.32**	5.16**	1.11	Baseline	.61	.59	.48
				Training	.62	.57	.50
				Posttraining	.64	.67	.52
Selection of respondents to questions: proportion student calls out answer.	2.32*	1.03	1.69*	Baseline	.06	.08	.07
				Training	.03	.07	.09
				Posttraining	.04	.05	.09
Quality of child's answers: proportion correct.	2.09*	2.66**	0.86	Baseline	.68	.71	.61
				Training	.66	.67	.63
				Posttraining	.68	.72	.66
Quality of child's answers: proportion wrong.	2.34*	24.07**	1.64*	Baseline	.04	.04	.03
				Training	.03	.05	.03
				Posttraining	.07	.07	.05
Quality of child's answers: proportion don't know and no response.	.72	.79	.83	Baseline	.04	.04	.03
				Training	.04	.04	.04
				Posttraining	.05	.04	.04
Teacher reaction to wrong answer: criticism.	.86	3.29**	1.25	Baseline	.001	.001	.000
				Training	.000	.000	.002
				Posttraining	.000	.008	.008
Teacher reaction to wrong answer: rephrases or gives clues.	3.79**	23.22**	1.40	Baseline	.07	.06	.03
				Training	.06	.08	.05
				Posttraining	.13	.14	.07
Teacher reaction to "I don't know" or no response: repeats question.	3.57**	21.14**	0.89	Baseline	.03	.05	.02
				Training	.08	.07	.03
				Posttraining	.11	.11	.08

TABLE 8 (continued)

Repeated Measures Means and F-Ratios for Discrete Process Variables: Baseline, Training, and Posttraining Phases

Variable	F-ratios			Means			
	F_{Trmt}	$F_{Time\ of\ Yr}$	$F_{Trmt \times Time}$	Time of Year	Maximal	Minimal	Control
Self and opinion questions: self questions ÷ process + product + choice questions.	0.78	2.63**	2.52**	Baseline	.21	.18	.11
				Training	.12	.13	.15
				Posttraining	.16	.10	.11
Self and opinion questions: opinion questions ÷ process + product + choice questions.	2.15*	0.78	0.18	Baseline	.008	.013	.001
				Training	.003	.007	.001
				Posttraining	.005	.007	.001
Self and opinion questions: proportion student opinions accepted.	0.63	0.92	0.32	Baseline	.002	.004	.002
				Training	.003	.003	.000
				Posttraining	.004	.005	.003
Combined teacher feedback data: repeat + rephrase or gives clue ÷ repeat + rephrase or gives clue + new question.	2.11*	19.89**	1.03	Baseline	.08	.08	.06
				Training	.12	.12	.07
				Posttraining	.18	.16	.10
Supervising (teacher activities).	0.61	6.90**	1.80*	Baseline	.05	.06	.06
				Training	.07	.07	.05
				Posttraining	.07	.08	.07
Instruction (teacher responses).	0.08	3.55**	0.74	Baseline	.15	.14	.16
				Training	.13	.14	.13
				Posttraining	.13	.13	.13
Adult commands, requests, and directs questions to individual children, non-academic.	0.14	9.49**	3.21**	Baseline	.013	.013	.009
				Training	.010	.011	.014
				Posttraining	.010	.007	.007

TABLE 8 (continued)

Repeated Measures Means and F-Ratios for Discrete Process Variables: Baseline, Training, and Posttraining Phases

Variable	F-ratios			Means			
	F_{Trmt}	$F_{Time\ of\ Yr}$	$F_{Trmt \times Time}$	Time of Year	Maximal	Minimal	Control
Adult commands, requests, and directs questions to groups of children, academic.	0.22	2.30*	1.29	Baseline	.018	.018	.016
				Training	.011	.015	.019
				Posttraining	.013	.012	.016
Adult academic commands, requests, and directs questions to children.	1.77*	2.36*	1.70*	Baseline	.06	.06	.04
				Training	.05	.05	.05
				Posttraining	.05	.05	.04
Child response, non-academic	0.54	7.29**	3.54**	Baseline	.05	.06	.05
				Training	.04	.05	.06
				Posttraining	.05	.04	.05
Adult instruction, academic.	0.08	3.54**	0.74	Baseline	.15	.14	.16
				Training	.13	.14	.13
				Posttraining	.13	.13	.13
Child to adult, all verbal except response.	1.82*	2.13*	0.86	Baseline	.07	.05	.06
				Training	.06	.05	.05
				Posttraining	.06	.05	.06
Adult communication or attention focus, large group.	0.18	4.76**	2.34**	Baseline	.21	.21	.24
				Training	.18	.21	.17
				Posttraining	.19	.19	.19
Adult attention to large group.	0.16	3.15**	1.59*	Baseline	.03	.02	.03
				Training	.04	.03	.03
				Posttraining	.03	.04	.04

TABLE 8 (continued)

Repeated Measures Means and F-Ratios for Discrete Process Variables: Baseline, Training, and Posttraining Phases

Variable	F-ratios			Means			
	F_{Trmt}	$F_{Time\ of\ Yr}$	$F_{Trmt \times Time}$	Time of Year	Maximal	Minimal	Control
Adult movement.	2.06*	6.47**	1.17	Baseline	.08	.06	.07
				Training	.07	.06	.07
				Posttraining	.08	.08	.09

** $p \leq .10$

* $.10 < p \leq .20$

TABLE 9

Summary of "More" and "Less" Academically-Oriented Ratings by Class

Teacher ID #	N of Students Rated Less Academically-Oriented	N of Students Rated More Academically-Oriented	% More Academically-Oriented
02	8	14	64
03	5	19	79
04	11	16	59
06	2	18	90
07	6	14	70
12	7	15	68
13	1	12	92
14	8	17	68
15	4	14	78
16	3	17	85
17	6	18	75
18	5	21	81
21	6	11	65
22	4	13	76
23	10	12	55
24	7	12	63
25	7	8	53
28	12	6	33
29	7	15	68
32	10	8	45
33	13	7	35
35	6	8	57
37	0	19	100

TABLE 10

Correlations between Process Variables and the Measure "Percentage of Students More Academically-Oriented" for Training Group Classes (N = 23)

Variable	Correlations (Pearson rs) Total Sample (N = 23)	District I (N = 12)	District II (N = 11)
High Inference Variables			
Amount of Personal, Private Interactions, A.M.	-.28	-.08	-.35
Positive Affect, A.M.	-.17	-.35	-.12
Negative Affect, A.M.	-.11	.05	.06
Pupil Passivity, A.M.	-.06	.11	.16
Amount of Personal Private Interaction, P.M.	-.27	.12	-.51
Positive Affect, P.M.	-.09	-.41	-.05
Negative Affect, P.M.	-.23	-.02	-.25
Pupil Passivity, P.M.	-.23	.12	-.09
Low Inference Variables			
Teacher rephrases or gives clue after wrong answer, during training.	.29	.20	.21
Teacher rephrases or gives clue after wrong answer, during post-training.	.24	.21	.47
Student "call outs", during training.	-.04	-.18	.07
Student "call outs", during posttraining.	-.07	.22	.04
Ratio of "self-related" questions to academic questions, during training.	-.49	-.03	-.50
Ratio of "self-related" questions to academic questions, during post-training.	-.41	-.20	-.49

TABLE 11

Correlations Between Process Measures and Adjusted Achievement: Comparisons of rs Across Studies

Variable #[1]	r in Prior Study	r in Present Study--N = 33 Observations based on Whole Year		
Behavior Management:		Vocabulary	Reading Comprehension	Total Reading
*BM01	+.35[2]	+.26	-.08	+.13
BM02	-.62	+.20	-.04	+.12
*BM03	+.41 LoSES; -.35 HiSES	+.08	-.22	-.09
BM04	+.45	-.13	-.11	-.14
BM05	+.43	-.45	-.05	-.33
BM06	+.35	-.19	-.08	-.18
*BM07	-.31	-.14	-.10	-.15
*BM08	-.21[3]	+.13	-.21	-.06
*BM09	+[4]	+.22	+.14	+.26
*BM10	+[4]	+.20	+.09	+.22
BM11	-[4]	+.08	+.05	+.07
Instructional Methods:				
*IM01	-.25 HiSES	-.12	-.17	-.19
IM02	-.36	+.22	-.20	+.02
IM03	-.36	+.26	-.19	+.06
*IM04	-.29	-.12	-.17	-.19
IM05	-.32	-.10	+.04	-.03
*IM06	-.32	+.21	-.30	-.05
IM07	-.39	+.16	-.03	+.10
IM08	+.46	+.07	-.14	-.06
IM09	+.28	-.03	+.12	+.04
IM11	-.29	+.32	-.11	+.17
IM12	+.32	-.08	+.03	-.04
IM14	-.28	+.03	-.13	-.05
*IM15	+.29	+.10	+.07	+.10
IM18	-.36	+.31	-.28	+.04
IM19	+.48	+.005	-.02	-.01
*IM20	+.13	+.23	-.22	+.04
*IM21	-.26	-.21	+.05	-.13
*IM23	+.27	+.10	+.07	+.10
IM24	+.34	(No correlation coefficients computed due to no variance for these Curriculum Survey variables.)		
IM25	+.48			
*IM26	+.28	+.33	+.01	+.25
IM27	-.31	+.06	-.13	+.13
IM28	+.19	-.16	+.15	-.03
IM29	+.31	+.10	-.006	+.06
IM30	+.45[5]	-.02	-.03	-.04
*IM31	-.46	-.35	+.18	-.14
Questioning and Feedback:				
*QF01	+.55	+.37	-.10	+.19
*QF02	-.32	-.32	+.05	-.18

TABLE 11 (continued)

Correlations Between Process Measures and Adjusted Achievement: Comparisons of rs Across Studies

Variable #[1]	r in Prior Study	r in Present Study--N = 33 Observations based on Whole Year: Vocabulary	Reading Comprehension	Total Reading
Questioning and Feedback (cont'd)				
QF03	+.51 LoSES; -.31 HiSES	+.10	-.06	+.01
*QF04	-.17 LoSES: +.39 HiSES	+.30	+.11	+.29
*QF05	-.63 LoSES; +.41 HiSES	+.02	+.003	+.04
*QF07	r.45	+.06	+.18	+.17
*QF08	-.15 LoSES; +.61 HiSES	-.06	+.26	+.13
QF09	+.65 LoSES; -.19 HiSES	+.31	-.04	+.19
QF10	**[6]	+.15	+.008	+.11
*QF11	-.19 LoSES; +.63 HiSES	-.06	+.26	+.13
QF13	-.49	+.11	+.10	+.16
QF19	**[6]	+.22	+.02	+.17
*QF21	-.26	-.06	-.26	-.21

"*"to the left of the variable # indicates some evidence of confirmation of prior results, according to the criteria described in text: two of the three rs with achievement in this study agree in direction with the earlier study's correlation, with one of the two rs greater than .15, or all three rs in this study agree in sign, with no stipulation about the magnitude of the rs.

Footnotes

[1] See Table 1 for definitions of variables.

[2] In the Brophy-Evertson study, "use of a system of rules" was measured with two rating scales. The r of +.35 is the average of the 2 process-product correlations (after changing the direction of one, since it was worded oppositely).

[3] The -.21 correlation was for a variable representing "one or more disciplinary errors," including target, timing, and overreactions.

[4] BM09, BM10, and BM11 were not selected for the teacher training program on the basis of prior correlations, but because of their usefulness as explanatory concepts; the "+" or "-" indicates the expected direction of correlation considering the nature of the recommendations.

[5] The direction of the correlation from the prior study was changed (from - to +) to indicate agreement with the name of the variable--"structured" teacher activity--as was measured in this study, i.e., this is the same as a negative r for "pupil free choice" (the process measure in the prior study was a bipolar factor).

[6] The process variables forming "sustaining feedback" and "ratio of rephrase to other feedback" showed differential correlations between SES subsamples and suggested different patterns according to whether the teacher rephrased or repeated the question. The coding system in this study did not distinguish rephrasing and repeating, so it is not possible to assess confirmation of prior results for these two variables.

Appendix A

Means and Standard Deviations of Implementation Components[1]

[1]The variable codes refer to names of implementation components in Table 4. The group codes in these breakdowns are: 1 = Maximal training, 2 = Minimal training, and 3 = Control group. Means are given for treatment groups separately by district; main effect means (collapsing across districts) can be obtained by adding sums for district 1 and district 2 and then dividing by the appropriate N -- 12 for the maximal training group, 11 for the minimal training group, and 10 for the control group.

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE BM1
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			-0.0017	-0.0001	1.9478	3.7940	(33)
DIST	1.		0.4605	0.0256	1.7059	2.9100	(18)
TRMT	1.		2.1762	0.3627	2.0250	4.1006	(6)
TRMT	2.		4.9863	0.8310	0.8559	0.7325	(6)
TRMT	3.		-6.7020	-1.1170	1.6061	2.5795	(6)
DIST	2.		-0.4622	-0.0308	2.2664	5.1366	(15)
TRMT	1.		-2.5302	-0.4217	2.5590	6.5464	(6)
TRMT	2.		5.1986	1.0397	1.0775	1.1610	(5)
TRMT	3.		-3.1306	-0.7826	2.9223	8.5398	(4)

TOTAL CASES = 33

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE BM2
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			-0.0013	-0.0000	1.9111	3.6524	(33)
DIST	1.		-5.1213	-0.2845	1.9771	3.9089	(18)
TRMT	1.		-0.7156	-0.1193	2.3407	5.4787	(6)
TRMT	2.		4.2798	0.7133	0.8078	0.6526	(6)
TRMT	3.		-8.6855	-1.4476	2.0756	4.3083	(6)
DIST	2.		5.1200	0.3413	1.8365	3.3729	(15)
TRMT	1.		-0.6307	-0.1051	1.6705	2.7905	(6)
TRMT	2.		5.8589	1.1718	1.5116	2.2849	(5)
TRMT	3.		-0.1082	-0.0270	2.5127	6.3139	(4)

TOTAL CASES = 33

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE BM3
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE		N
FOR ENTIRE POPULATION			-0.0004	-0.0000	1.8843	3.5507	(	33)
DIST	1.		11.8525	0.6585	0.9893	0.9787	(	18)
TRMT	1.		0.0021	0.0003	1.3872	1.9243	(	6)
TRMT	2.		7.1181	1.1863	0.4118	0.1696	(	6)
TRMT	3.		4.7323	0.7887	0.5995	0.3594	(	6)
DIST	2.		-11.8529	-0.7902	2.3876	5.7005	(	15)
TRMT	1.		2.5436	0.4239	1.0417	1.0851	(	6)
TRMT	2.		-9.5087	-1.9017	2.6936	7.2554	(	5)
TRMT	3.		-4.8878	-1.2219	3.1409	9.8656	(	4)

TOTAL CASES = 33

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE BM4
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE		N
FOR ENTIRE POPULATION			0.0118	0.0004	1.6802	2.8229	(	33)
DIST	1.		10.7684	0.5982	1.3610	1.8522	(	18)
TRMT	1.		0.0383	0.0064	1.2676	1.6560	(	6)
TRMT	2.		6.7548	1.1275	1.0111	1.0222	(	6)
TRMT	3.		3.9653	0.6609	1.6900	2.8561	(	6)
DIST	2.		-10.7566	-0.7171	1.7867	3.1922	(	15)
TRMT	1.		-6.3473	-1.0579	2.5794	6.6531	(	6)
TRMT	2.		-0.8838	-0.1768	1.1690	1.3666	(	5)
TRMT	3.		-3.5255	-0.8814	1.1096	1.2313	(	4)

TOTAL CASES = 33

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE COMPSMS
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE		N
FOR ENTIRE POPULATION			-0.0072	-0.0002	3.8119	14.5309	(	33)
DIST	1.		-5.7924	-0.3218	4.0714	16.5765	(	18)
TRMT	1.		4.5797	0.7633	5.1230	26.2447	(	6)
TRMT	2.		4.6572	0.7762	2.5372	6.4374	(	6)
TRMT	3.		-15.0293	-2.5049	3.8858	15.0994	(	6)
DIST	2.		5.7852	0.3857	3.5766	12.7924	(	15)
TRMT	1.		0.2262	0.0377	2.2013	4.8456	(	6)
TRMT	2.		12.9122	2.5824	3.4462	11.8760	(	5)
TRMT	3.		-7.3532	-1.8383	4.5724	20.9072	(	4)

TOTAL CASES = 33

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE COMPBMS
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE		N
FOR ENTIRE POPULATION			-0.0626	-0.0019	1.2583	1.5832	(	33)
DIST	1.		2.1612	0.1201	1.1084	1.2285	(	18)
TRMT	1.		2.6009	0.4668	1.0801	1.1666	(	6)
TRMT	2.		0.1973	0.0329	1.2795	1.6371	(	6)
TRMT	3.		-0.8369	-0.1395	1.0672	1.1390	(	6)
DIST	2.		-2.2239	-0.1483	1.4439	2.0850	(	15)
TRMT	1.		-1.8563	-0.3094	1.2794	1.6369	(	6)
TRMT	2.		-0.0475	-1.2095	0.4859	0.2361	(	5)
TRMT	3.		5.8799	1.4200	1.2160	1.4788	(	4)

TOTAL CASES = 33

- D E S C R I P T I O N O F

CRITERION VARIABLE COMPIM1
BROKEN DOWN BY DIST
BY TRMT

- -

VARIABLE	CODE	VALUE LABEL
FOR ENTIRE POPULATION		
DIST	1.	
TRMT	1.	
TRMT	2.	
TRMT	3.	
DIST	2.	
TRMT	1.	
TRMT	2.	
TRMT	3.	

TOTAL CASES = 33

- D E S C R I P T I O N O

CRITERION VARIABLE COMPIM2
BROKEN DOWN BY DIST
BY TRMT

- -

VARIABLE	CODE	VALUE LABEL
FOR ENTIRE POPULATION		
DIST	1.	
TRMT	1.	
TRMT	2.	
TRMT	3.	
DIST	2.	
TRMT	1.	
TRMT	2.	
TRMT	3.	

TOTAL CASES = 33

SUBPOPULATIONS

SUM	MEAN	STD DEV	VARIANCE		N
-0.0039	-0.0001	1.0269	1.0587	(	33)
-0.6097	-0.0339	1.1315	1.2802	(	18)
0.9389	0.1565	0.6046	0.3655	(	6)
3.4853	0.5815	1.2045	1.4509	(	6)
-5.0379	-0.8397	1.1221	1.2591	(	6)
0.6058	0.0404	0.9285	0.8621	(	15)
0.0066	0.0011	0.7851	0.6163	(	6)
-1.2364	-0.2473	0.9781	0.9566	(	5)
1.8356	0.4589	1.1600	1.3457	(	4)

F SUBPOPULATIONS

SUM	MEAN	STD DEV	VARIANCE		N
-0.0043	-0.0001	2.7977	7.8269	(	33)
1.1198	0.0622	2.6744	8.2622	(	18)
-1.0371	-0.1728	3.6286	13.1669	(	6)
11.5176	1.9196	2.3980	5.7502	(	6)
-9.3608	-1.5601	1.3452	1.8097	(	6)
-1.1241	-0.0749	2.8012	7.8465	(	15)
0.3622	0.0604	1.7533	3.0739	(	6)
6.9913	1.3963	3.8350	14.7073	(	5)
-8.4777	-2.1194	1.6300	2.6569	(	4)

- - - -

SUBPOPULATIONS ----------

SUM	MEAN	STD DEV	VARIANCE		N
-0.0409	-0.0012	2.2860	5.2257	(	33)
11.9161	0.6620	1.7320	2.9998	(	18)
2.3365	0.3894	2.1057	4.4336	(	6)
9.6501	1.6084	1.4733	2.1707	(	6)
-0.0705	-0.0118	1.3734	1.8862	(	6)
-11.9570	-0.7971	2.6566	7.0575	(	15)
-2.4056	-0.4009	2.8443	8.0902	(	6)
-5.1467	-1.0297	2.7652	7.6462	(	5)
-4.4027	-1.1007	2.9546	8.7294	(	4)

SUBPOPULATIONS ----------

SUM	MEAN	STD DEV	VARIANCE		N
0.0008	0.0000	1.7404	3.0291	(	33)
2.9029	0.1613	1.5298	2.3404	(	18)
1.1966	0.1994	1.6708	2.7915	(	6)
2.1168	0.3531	1.4019	1.9652	(	6)
-0.4125	-0.0688	1.7582	3.0911	(	6)
-2.9021	-0.1935	2.0021	4.0062	(	15)
1.3862	0.2310	1.9027	3.6201	(	6)
2.0407	0.4081	2.5141	6.3209	(	5)
-0.3289	-1.5822	0.8417	0.7085	(	4)

DESCRIPTION OF

CRITERION VARIABLE COMPIMA
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL
FOR ENTIRE POPULATION		
DIST	1.	
TRMT	1.	
TRMT	2.	
TRMT	3.	
DIST	2.	
TRMT	1.	
TRMT	2.	
TRMT	3.	

TOTAL CASES = 33

DESCRIPTION OF

CRITERION VARIABLE COMPIMA
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL
FOR ENTIRE POPULATION		
DIST	1.	
TRMT	1.	
TRMT	2.	
TRMT	3.	
DIST	2.	
TRMT	1.	
TRMT	2.	
TRMT	3.	

TOTAL CASES = 33

SUBPOPULATIONS ----------

SUM	MEAN	STD DEV	VARIANCE	N
0.1901	0.0058	2.0082	4.0329	(33)
1.8119	0.1007	1.9383	3.7571	(18)
-3.9637	-0.6606	2.0256	4.1032	(6)
-1.4976	-0.2496	1.1491	1.3203	(6)
7.2733	1.2122	2.2417	5.0253	(6)
-1.6218	-0.1081	2.1518	4.6304	(15)
-0.0343	-0.0057	1.8007	3.2425	(6)
2.6086	0.5217	3.0318	9.1920	(5)
-4.1961	-1.0490	1.4442	2.0858	(4)

SUBPOPULATIONS ----------

SUM	MEAN	STD DEV	VARIANCE	N
0.3360	0.0102	3.4561	11.9448	(33)
1.8388	-0.6577	3.4710	12.0482	(18)
6.2883	-1.0480	5.0796	25.8023	(6)
2.5975	0.4329	2.2542	5.0816	(6)
8.1480	-1.3580	2.6073	7.6612	(6)
2.1748	0.8117	3.3760	11.4108	(15)
2.5046	-0.4174	4.7668	22.7222	(6)
6.8561	1.3716	1.9811	3.9247	(5)
7.9213	1.9553	2.2044	4.8593	(4)

DESCRIPTION OF

CRITERION VARIABLE QF1
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL
FOR ENTIRE POPULATION		
DIST	1.	
TRMT	1.	
TRMT	2.	
TRMT	3.	
DIST	2.	
TRMT	1.	
TRMT	2.	
TRMT	3.	

TOTAL CASES = 33

DESCRIPTION OF

CRITERION VARIABLE QF2
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL
FOR ENTIRE POPULATION		
DIST	1.	
TRMT	1.	
TRMT	2.	
TRMT	3.	
DIST	2.	
TRMT	1.	
TRMT	2.	
TRMT	3.	

TOTAL CASES = 33

SUBPOPULATIONS - - - - - - - - - - - - - - - - - -

SUM	MEAN	STD DEV	VARIANCE		N
0.0001	0.0000	1.0001	1.0003	(	33)
1.7406	0.0967	0.7226	0.5221	(	18)
1.5220	0.2537	0.6275	0.3938	(	6)
0.8867	0.1481	0.8860	0.7850	(	6)
-0.6701	-0.1117	0.7151	0.5114	(	6)
-1.7405	-0.1160	1.2751	1.6260	(	15)
2.0803	0.3467	0.9033	0.8160	(	6)
2.2918	0.4584	0.7715	0.5952	(	5)
-6.1127	-1.5282	1.3406	1.7972	(	4)

SUBPOPULATIONS - - - - - - - - - - - - - - - - - -

SUM	MEAN	STD DEV	VARIANCE		N
+0.0277	+0.0008	1.0010	1.0019	(	33)
+1.0998	+0.0611	1.1457	1.3127	(	18)
+3.1127	+0.5188	0.6354	0.4036	(	6)
+2.0428	+0.3405	0.8700	0.7569	(	6)
-4.0557	-0.6760	1.5184	2.3056	(	6)
-1.0721	-0.0715	0.8281	0.6856	(	15)
+2.6429	+0.4405	0.6183	0.3823	(	6)
-1.3765	-0.2741	0.7679	0.5897	(	5)
-2.3445	-0.5861	0.9118	0.8314	(	4)

DESCRIPTION O

CRITERION VARIABLE QF3
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL
FOR ENTIRE POPULATION		
DIST	1.	
TRMT	1.	
TRMT	2.	
TRMT	3.	
DIST	2.	
TRMT	1.	
TRMT	2.	
TRMT	3.	

TOTAL CASES = 33

DESCRIPTION O

CRITERION VARIABLE QF4
BROKEN DOWN BY DIST
BY TRMT

VARIABLE	CODE	VALUE LABEL
FOR ENTIRE POPULATION		
DIST	1.	
TRMT	1.	
TRMT	2.	
TRMT	3.	
DIST	2.	
TRMT	1.	
TRMT	2.	
TRMT	3.	

TOTAL CASES = 33

F SUBPOPULATIONS ----------

SUM	MEAN	STD DEV	VARIANCE		N
0.0068	0.0002	0.9979	0.9958	(	33)
-4.1820	-0.2323	0.6383	0.4074	(	18)
-2.4287	-0.4048	0.4809	0.2312	(	6)
-0.4538	-0.0756	0.8497	0.7219	(	6)
-1.2995	-0.2166	0.6055	0.3666	(	6)
4.1888	0.2793	1.2761	1.6284	(	15)
0.7463	0.1244	1.3123	1.7221	(	6)
5.3927	1.0785	1.3294	1.7668	(	5)
-1.9501	-0.4875	0.6887	0.4743	(	4)

F SUBPOPULATIONS ----------

SUM	MEAN	STD DEV	VARIANCE		N
0.0067	0.0003	1.7558	3.0828	(	33)
-1.2468	-0.0693	1.5657	2.4515	(	18)
-5.2178	-0.8696	0.7484	0.5601	(	6)
3.0547	0.5091	2.0537	4.2179	(	6)
0.9163	0.1527	1.5258	2.3279	(	6)
1.2555	0.0837	2.0139	4.0559	(	15)
-1.2324	-0.2054	1.7347	3.0091	(	6)
0.8265	0.1653	2.4041	5.7797	(	5)
1.6614	0.4154	2.4251	5.8813	(	4)

Teacher Self-Report in an Experiment on Teacher Education and Effectiveness

Alexis L. Mitman

Stanford University

School of Education

Center for Educational Research at Stanford

Program on Teaching Effectiveness

Abstract

Descriptive information about the 33 teachers participating in the experiment on teacher training and teacher effectiveness is summarized. The two major sources of information were a semi-structured interview and an extensive questionnaire. Results indicate the limited utility of teacher reports about changes in their own behavior when these changes are the focus of a training program. Other teacher self-reports on their general teaching practices, classroom environments, and students were found to be fairly accurate. Teacher descriptions of their reactions to the training and factors affecting its implementation are shown to be valuable in the evaluation and interpretation of other findings.

Teacher Self-Report in an Experiment on Teacher Education and Effectiveness

This paper discusses the major findings from the teacher self-report data collected in the Program on Teaching Effectiveness third grade experiment on teacher training and effectiveness. These data were gathered from two sources: an interview and a questionnaire. First, an extensive interview was conducted with all 33 participating teachers at midyear. This interview was semi-structured and designed to elicit information on teacher attitudes toward the teacher training program and their classroom situations, and on the salience and implementation of the training recommendations. Different versions of the interview were developed to accommodate the different treatment groups. Interviews for "maximally" and "minimally" trained teachers took approximately one hour. Control-group teachers were given a shorter interview which did not deal with the training variables. The interviews were administered by program staff members, all of whom had received training on the content of the interview and on interviewing techniques. The teacher interview was the largest single source of descriptive information. It was felt that collecting records of this more complex and idiographic nature, often termed "qualitative," would provide a valuable interpretive and evaluative supplement to the quantitative data from the experiment.

The second source of teacher self-report data, a teacher questionnaire, was completed by all teachers at the beginning and end of the school year. This questionnaire included measures of teacher personality traits, teacher verbal ability, classroom environment, teaching practices and teaching concerns.

Of central importance is the question of the validity of teacher self-report data. In other words, to what extent can these kinds of data be accepted as accurate indicators of actual teacher behavior and attitudes? Also important is the question of the general usefulness of the complex idiographic teacher reports which the interview yielded. The self-report data from this study were examined with regard to four major areas: (1) teachers' reports of changes in their training-related teaching, (2) (2) teachers' reports about their general teaching practices, classroom environments, and students, (3) teachers' reports about the strengths and weaknesses of the training procedures, and (4) teachers' reports on the uncontrolled factors that may have influenced implementation of the training.

Teachers' Reports on Their Training-Related Teaching Behavior

In both the questionnaire and interview, teachers were asked to describe changes in their use of the teaching practices addressed in the training packets. These descriptions were analyzed in terms of whether the relative changes in teaching behavior reported by three different training groups could be substantiated by the low inference observation data which is reported in detail in the Stayrook and Crawford (1978) paper.

The teacher questionnaire. Given at the beginning and end of the year, the teacher questionnaire included self-report measures of classroom environment, teaching practices, and teaching concerns. All pre-post correlations between these measures, with the exception of those measures dealing with teaching practices specifically addressed in the training recommendation packets, showed strong positive relationships beyond the .01 level of significance. (See Table 1.) The correlations between the pre- and post-reports on the degree to which recommended practices in each of the three training areas (behavior management, instructional methods, and questioning and

Table 1

Means, Standard Deviations, and Intercorrelations of Teacher Pre- and Post-Questionnaire Variables

QUESTIONNAIRE VARIABLE	Pretest Means and Standard Deviations	Posttest Means and Standard Deviations	r	p value
Structuredness of Teaching Style, Stallings and Kaskowitz, 1974	$\bar{X}$ = 26.313 SD = 5.221	$\bar{X}$ = 31.500 SD = 6.900	.61	.001
Teacher Self-Rating of Use of Motivational Techniques	$\bar{X}$ = 31.656 SD = 5.350	$\bar{X}$ = 29.606 SD = 6.750	.74	.001
Teaching Practices Subscale-Behavior Management	$\bar{X}$ = 13.273 SD = 1.567	$\bar{X}$ = 14.606 SD = 1.836	.22	.106
Teaching Practices Subscale-Instructional Methods	$\bar{X}$ = 30.485 SD = 4.154	$\bar{X}$ = 38.939 SD = 5.099	.18	.152
Teaching Practices Subscale-Questioning and Feedback	$\bar{X}$ = 41.697 SD = 4.433	$\bar{X}$ = 51.091 SD = 8.308	.07	.356
Teaching Practices--Total Score	$\bar{X}$ = 82.424 SD = 15.008	$\bar{X}$ = 104.636 SD = 13.463	-.15	.202
Dimensions of Schooling--Total Score[1]	$\bar{X}$ = 90.156 SD = 14.014	$\bar{X}$ = 22.485 SD = 4.678	.56	.001
Teacher Concern with Doing the Job Well and Being Liked	$\bar{X}$ = 16.969 SD = 3.873	$\bar{X}$ = 17.152 SD = 3.930	.44	.006
Teacher Concern with Getting Along with Children and School Personnel	$\bar{X}$ = 20.188 SD = 5.981	$\bar{X}$ = 16.152 SD = 5.821	.58	.001
Teacher Concern with Providing Individualized and Reality-Based Instruction	$\bar{X}$ = 11.625 SD = 2.366	$\bar{X}$ = 11.061 SD = 2.633	.47	.004
Teacher Concern with Guiding Students and Providing Stable Emotional and Intellectual Climate	$\bar{X}$ = 28.031 SD = 6.723	$\bar{X}$ = 26.909 SD = 8.001	.46	.004
Teacher Concern with Physical Limitations in Terms of Time and Materials	$\bar{X}$ = 9.344 SD = 4.051	$\bar{X}$ = 8.364 SD = 3.831	.57	.001
Teacher Concern with Being Favorably Evaluated for Doing a Good Job	$\bar{X}$ = 9.688 SD = 2.978	$\bar{X}$ = 8.364 SD = 3.380	.59	.001
Teacher Satisfaction with Job	$\bar{X}$ = 15.125 SD = 4.696	$\bar{X}$ = 13.849 SD = 4.797	.55	.001
Teacher Aspirations	$\bar{X}$ = 12.469 SD = 3.742	$\bar{X}$ = 11.424 SD = 3.873	.60	.001

[1]Posttest scores on this instrument were reduced by a factor of 4.

feedback) were used did not differ significantly from zero. The correlation between the pre- and post-report on the degree to which the total group of recommended practices were used also was nonsignificant ($r = -.150$, $p = .202$). Together, these correlations indicate what we should expect if the training was effective, namely, that the reported use of training practices for some teachers changed substantially from the beginning to the end of the school year. Treatment group means for the post-questionnaire measures indicated that the maximal and minimal teachers reported a dramatic increase in their use of the training recommendations in all three areas (relative to the no-training teachers)--particularly the questioning and feedback recommendations.

Low inference observational results on teacher implementation of the training recommendations suggest that these self-reported changes have limited validity. (See Stayrook & Crawford, 1978, for a detailed presentation.) For example, the observational findings indicated that maximal and minimal teachers implemented the component behaviors of the behavior management and questioning and feedback training areas more consistently than the component behaviors of the instructional methods training area.

The teacher interview. In the interview, maximal and minimal teachers were asked a number of direct and indirect questions about implementation of the training recommendations. Although all the interview questions allowed for open-ended responses, the pattern of answers to a small number of questions was amenable to simple classification (usually in terms of two or three levels). Thus, 15 "quantifiable" teacher interview variables were created. A list of these variables appears in Table 2. It should be emphasized that these variables summarize only a small fraction of the interview data. The great majority of these data remain in the form of prose descriptions for each individual teacher. The major purpose of the

Table 2

Variables from the Teacher Interview

VARIABLE NAME	MEANING AND SCORING	TEACHERS TO WHOM QUESTION APPLIES	CORRESPONDING INTERVIEW QUESTION
YRSINSCH	Length of time (yrs.) teaching 3rd graders at this school	All	3
INTV01	Academic standing of this class relative to ones in past: 3(better) 2(same) 1(worse)	All	5
INTV02	Classroom architecture: 2(closed) 1(open)	All	6
INTV03	Frequency of student assigned-seating: 3(always) 2(sometimes) 1(never)	All	7-9
INTV04	Desire for additional help beyond packets: 2(yes) 1(no)	Minimum teachers only	10
INTV05	Time spent reading each of the training packets: 3(>2 hrs.) 2(1-2 hrs.) 1(<1 hr.)	Maximum and Minimum Teachers only	12
INTV06	Severity of typical classroom discipline problems: 3(high) 2(medium) 1(low)	"	21
INTV07	Difficulties with classroom management recommendations: 2(some) 1(none)	"	28
INTV08	Difficulties with instructional methods recommendations: 2(some) 1(none)	"	40
INTV09	Difficulties with question-feedback recommendations: 2(some) 1(none)	"	53
INTV10	Degree of sophistication in describing general purpose of training: 3(high) 2(medium) 1(low)	"	56
INTV11	Academic composition of class: 3(mostly More Academically Oriented) 2(half and half) 1(mostly Less Academically Oriented)	"	71
INTV12	Estimated increase in academic time [illegible]e: 3(a lot (25% +)) 2(medium (15-20%)) 1(none/small (5-10%))	"	78-79
INTV13	Degree of accord between teacher's own philosophy and training philosophy: 3(great) 2(sometimes) 1(not much)	"	95
SALIENCE	Total number of training recommendations mentioned during the interview (22 possible)	"	23, 35, 48

interviews would be defeated if an attempt had been made to quantify responses to all of the questions.

It was hypothesized that some of the interview variables would be associated with the observational data on implementation--specifically, the variables reflecting teachers' reported difficulties in implementing recommendations in each of the three training areas (see variables INTV07-INTV09), their ability to summarize the interview recommendations (SALIENCE) and to describe the purpose of the training (INTV10), and their reported degree of philosophical agreement with the training recommendations (INTV13). When the interview variables were correlated with training and posttraining observational implementation measures (three composite measures for each training area and one total score), six of the 60 correlations were significant at the $p < .05$ level. Table 3 presents the correlation matrix. Teacher reports of the relative academic standing of their class (INTV01), their classroom architecture (INTV02), and their use of assigned seating (INTV03) all were positively associated with training and posttraining total use of the instructional methods recommendations ($rs = .29$, .46, and .43, respectively). These results suggest that regardless of teacher training group assignment, classes rated as more academically productive, with closed-room architecture, and with assigned seating were observed making use of the instructional methods recommendations. INTV01 was postively associated with total implementation at the $p < .10$ level ($r = .24$).

Teachers who reported having some difficulties implementing the instructional methods recommendations (INTV08) had higher implementation scores in instructional methods than teachers reporting no difficulties ($r = .48$). Perhaps teachers who were involved in thinking about and

Table 3

Correlations Between Interview Variables and Composite Implementation Measures[1]

	Behavior Management	Instructional Methods	Questioning and Feedback	Total
INTV01 (Academic Standing of this Class Relative to Ones in Past)	r = .16	r = .29	r = .01	r = .24
	p = .184	p = .049	p = .471	p = .086
INTV02 (Classroom Architecture)	r = -.13	r = .46	r = .11	r = .19
	p = .234	p = .004	p = .271	p = .147
INTV03 (Frequency of Student Assigned Seating)	r = -.08	r = .43	r = .09	r = .20
	p = .337	p = .007	p = .314	p = .135
INTV04 (Desire for Additional Help Beyond Packets)	r = .05	r = -.25	r = -.07	r = -.19
	p = .448	p = .231	p = .419	p = .287
INTV05 (Time Spent Reading Each of the Training Packets)	r = -.10	r = -.26	r = -.51	r = -.32
	p = .318	p = .115	p = .007	p = .071
INTV06 (Severity of Typical Classroom Discipline Problems)	r = -.29	r = -.27	r = .03	r = -.31
	p = .089	p = .105	p = .449	p = .072
INTV07 (Difficulties with Classroom Management Recommendations)	r = -.17	r = -.22	r = -.29	r = -.30
	p = .228	p = .166	p = .102	p = .093
INTV08 (Difficulties with Instructional Methods Recommendations)	r = .04	r = .48	r = .09	r = .35
	p = .428	p = .016	p = .351	p = .065
INTV09 (Difficulties with Question-Feedback Recommendations)	r = .12	r = .21	r = -.00	r = .19
	p = .298	p = .169	p = .498	p = .193
INTV10 (Degree of Sophistication in Describing General Purpose of Training)	r = -.32	r = -.17	r = .33	r = -.21
	p = .072	p = .213	p = .060	p = .170
INTV11 (Academic Composition of Class)	r = .01	r = -.11	r = .06	r = -.05
	p = .476	p = .301	p = .398	p = .415
INTV12 (Estimated Increase in Academic Time Use)	r = .24	r = .10	r = -.38	r = .11
	p = .139	p = .319	p = .038	p = .303
INTV13 (Degree of Accord Between Teacher's Own Philosophy and Training Philosophy)	r = .17	r = -.28	r = .06	r = .06
	p = .220	p = .095	p = .387	p = .398
SALIENCE (Total Number of Training Recommendations Mentioned During the Interview)	r = -.19	r = .12	r = .01	r = -.03
	p = .190	p = .287	p = .490	p = .439
YRSINSCH (Length of Time (Yrs.) Teaching 3rd Graders at this School)	r = .04	r = .01	r = .01	r = .03
	p = .405	p = .480	p = .486	p = .432

[1]Composite scores were developed by summing the z-scores ($\bar{X} = 0$, SD = 1) for the individual behavior variables in each training content area.

trying to practice these particular recommendations were those who honestly admitted having some difficulties. Teachers who reported having "no difficulties" may have lacked specific knowledge about the strategies they were supposed to be implementing. The moderate positive correlation between INTV08 and the SALIENCE variable (r = .27, p = .126) gives support to this interpretation.

The remaining two correlations do not lend themselves to an obvious interpretation. Teachers' reports of the average amount of time they spent reading each of the training packets (INTV05) and their estimated increase in academic time (INTV12) were negatively correlated with observed total use of the questioning and feedback recommendations (rs = -.51 and -.38). None of the interview variables correlated meaningfully with residualized student achievement outcomes.

In sum, the teacher self-report data on training-related behavior practices appear to represent both actual changes in the frequencies of the training behaviors and a relatively strong response bias on the part of trained teachers. Further evidence of a response bias comes from teacher interview responses that were not summarized by the above variables. For instance, when teachers were asked which of the recommendations in each of the training areas they had been using since the training, over half of the teachers responded by saying that they used all of them or always had used the recommendations. The observational data point to considerable greater variation in implementation among teachers than their own reports forecast.

The most direct implication of these findings is that the kinds of self-report measures used in this study should not be relied on as the sole measures of change in targeted teacher behaviors. It is pos-

sible that other kinds of self-report instruments, administered more systematically and without an experimenter's presence, might produce more valid results. Also, as the remainder of this paper will demonstrate, the questionable trustworthiness of self-report measures of training-related classroom behavior by no means implies that self-report measures about other aspects of teachers and their classrooms should receive diminished attention.

Teachers' Reports About Their General Teaching Practices, Classroom Environments and Students

The questionnaire and interview self-report data on teachers' general teaching practices, classroom environments, and students support an impression that the teachers gave fairly consistent and accurate reports about many kinds of phenomena. Turning first to reports about general teaching practices and classroom environments, we find that the strong positive correlations in Table 1 between pre- and post-questionnaire measures of all (except training-related) behaviors indicate the stability of these measures. Also, the relationships among teacher interview variables and teacher pre-questionnaire variables were examined since some variables from both instruments tapped similar aspects of teaching practice and classroom environment. There were strong positive correlations among conceptually related variables. High scores on two pre-questionnaire composite measures of classroom structuredness (The Dimensions of Schooling Questionnaire and The Classroom Description Rating Instrument), which reflect a high degree of structuredness, were associated with interview descriptions of closed classroom architecture (r_s = .34 and .46), assigned seating (r_s = .30 and .47), and a smaller increase in academic

time (rs = -.56, -.32). The last two correlations may be interpreted to mean that teachers in more structured classrooms already were devoting much time to academic activities prior to the training.

Correlations between interview variables reflecting teacher descriptions of their students and raw student outcomes provided information on the accuracy of teachers' descriptions of their students. There was impressive evidence that teachers are accurate in estimating and reporting the academic standing of their class relative to the classes of other teachers. Four interview variables correlated strongly ($p < .05$) with pre- and posttest raw score composites of vocabulary and reading. (See Table 4.) Teacher reports of the academic composition of their class and teacher reports of the academic standing of their class relative to previous classes were positively correlated with raw achievement scores (rs = .58 and .30, respectively). Teacher reports of the severity of typical discipline problems and teacher reports of the number of years spent teaching third graders at their current school were negatively correlated with raw achievement scores (pretest rs = -.56 and -.39, respectively). The latter correlations may mean that lower achieving classes actually do present more discipline problems to their teacher and that the more experienced teachers are assigned to teach them.

The patterns of relationships discussed here point to the kinds of teacher self-report items that can be interpreted as satisfactory indicators of classroom phenomena. Again, none of these measures was directly related to the training behaviors, and thus there was little reason to expect response bias on the part of teachers.

Table 4

Correlations Between Selected Interview Variables and Student Raw Achievement Outcomes

	Class Means Pretest Reading Score	Class Means Posttest Reading Score
INTV01 (Academic Standing of this Class Relative to Ones in Past)	r = .30 p = .045	r = .33 p = .029
INTV06 (Severity of Typical Classroom Discipline Problems)	r = -.56 p = .003	r = -.55 p = .003
INTV11 (Academic Composition of Class)	r = .58 p = .002	r = .61 p = .001
YRSINSCH (Length of Time (Yrs.) Teaching 3rd Graders at this School)	r = -.39 p = .013	r = -.36 p = .021

Teachers' Reports on the Strengths and Weaknesses of the Training Program

The findings discussed in this section and the next are intended to illustrate the general usefulness of the descriptive data gathered from the teacher interview. These findings are summaries of teachers' open-ended responses to interview questions. Although it is possible that portions of this information could have been collected with a structured questionnaire, it is unlikely that such an instrument would have permitted the richness and complexity of description that the interview yielded.

The interview proved to be effective in prompting teacher evaluations of the strengths and weaknesses of the training program. Although the answers to evaluative questions about the training varied considerably over teachers, there was consensus on some major points. This information is valuable as a basis for future revisions of the training program. It should be useful to those with a general interest in teacher inservice training methods as well.

One set of results dealt with the degree to which certain parts of the training information were more "salient" than others. This was assessed by asking each teacher during the interview to summarize as many of the recommendations as she could for each of the three training areas (behavior management, instructional methods, and questioning and feedback). A summary of these results is presented in Table 5. These salience scores varied considerably among teachers, ranging from 60% to 14%. The highest percentage (47%) was in the area of behavior management, followed by the question and feedback area (34%), and then instructional methods (23%). The two most frequently mentioned recommendations were (a) to call

Table 5
Percentage of Trained Teachers Mentioning Each of the 22 Training Recommendations, Broken Down by Training Group and District

TRAINING PACKET RECOMMENDATION	MAXIMALLY TRAINED TEACHERS			MINIMALLY TRAINED TEACHERS			TOTAL
	Dist. 1 N = 6	Dist. 2 N = 6	Dist. 1 & 2 N = 12	Dist. 1 N = 6	Dist. 2 N = 5	Dist. 1 & 2 N = 11	N = 23
1. Procedural Rules	0%	17%	8%	17%	0%	9%	9%
2. Timing Errors	67%	100%	83%	67%	40%	55%	70%
3. Target Errors	67%	83%	75%	67%	60%	64%	70%
4. Overreacting	50%	50%	50%	67%	80%	73%	61%
5. Monitoring and Withitness	17%	17%	17%	50%	20%	36%	26%
6. Appropriate Seatwork	17%	0%	8%	17%	20%	18%	9%
7. Minimizing Directions	33%	67%	50%	50%	40%	46%	48%
8. Use of Large Groups	67%	50%	58%	67%	0%	36%	48%
9. Use of Materials	33%	0%	17%	33%	20%	27%	17%
10. Visual Demonstrations and Phonics Exercise	0%	50%	25%	0%	20%	9%	17%
11. General Use of Academic Questioning & Answering Sessions	0%	0%	0%	0%	0%	0%	0%
12. Call Student's Name First	83%	83%	83%	67%	60%	64%	74%
13. Discourage Volunteers and Call-Outs	33%	50%	42%	33%	60%	46%	44%
14. Discourage Pupil-Initiated Questions and Comments	0%	0%	0%	0%	0%	0%	0%
15. Difficulty Level of Questions	33%	17%	25%	67%	60%	64%	44%
16. Use of Praise and Criticism	17%	50%	33%	67%	0%	36%	35%
17. Getting Responses from Less Academically-Oriented	67%	100%	83%	67%	80%	73%	78%
18. Getting Responses from More Academically-Oriented	50%	67%	58%	33%	20%	27%	44%
19. Giving Answer for Partially Correct Answers	17%	17%	17%	17%	20%	18%	17%
20. Length of Feedback	17%	33%	25%	17%	20%	18%	22%
21. Use of Brief Fast-Paced Feedback in Reading Groups	33%	17%	25%	33%	0%	18%	22%
22. Occasional Use of "Why" Explanations	0%	0%	0%	0%	0%	0%	0%

the student's name first when asking a question (74%), and (b) to always get some response from less academically-oriented students (78%). (Some recommendations were differentiated according to the student's characteristic academic orientation which was defined for teachers in behavioral terms.) Both of these recommendations come from the training packet on questioning and feedback. There was evidence reported in the Stayrook and Crawford (1978) paper showing that these two components were successfully implemented by training group teachers. It seems that, while the overall "salience" score is not a strong predictor of total implementation, the training behaviors that were mentioned by most (3/4) of the teachers were successfully implemented.

Comparing maximally trained and minimally trained teachers' attitudes toward the training also was important, since one purpose of this experiment was to contrast the effectiveness of an expensive with an inexpensive form of inservice training. The interview asked maximally trained teachers to discuss the ways in which they felt the workshop sessions (given only to the maximally trained teachers) helped them understand and use the training recommendations.

The most frequent response to this question, given by over half of the maximally trained group teachers, was that the sessions gave them the opportunity to communicate with other teachers. More specifically, teachers mentioned getting to hear the examples and viewpoints of other teachers and receiving reassurances from other teachers that they were not alone in having problems--general teaching problems and problems with implementation. A second general kind of response from maximally trained teachers, mentioned by over one third of the group, was that the sessions provided some oral review of the readings and thereby reinforced learning

of the training content. In addition, maximally trained teachers frequently mentioned having the opportunity to ask questions and receive clarification. Similar kinds of responses resulted when maximal teachers were asked to describe the major strengths of the training experience. The opportunity to "share" and interact with other teachers and with PTE staff was the most frequently mentioned strength. Review and clarification of content and viewing the videotape demonstration also were mentioned as strengths.

There was little agreement about the worst aspects of the training sessions. Two teachers mentioned problems with personal dynamics (they complained that some teachers dominated the discussions) and inefficient use of time. There was mixed opinion about the usefulness of role-playing exercises and the convenience of the meeting-time arrangements. When the maximally trained teachers were asked to suggest ways in which the training sessions could be improved, a variety of responses was given. No one suggestion, however, was mentioned by more than three teachers. The most common suggestions were that the sessions include more videotape demonstrations, more realistic practice situations, more role-playing exercises, and, on the other hand, that role playing be eliminated altogether.

In short, the predominant impression from the maximally trained teachers was that the training sessions were a very beneficial component of the training program. Most teachers expressed the opinion that having the chance to meet regularly to discuss the training facilitated understanding and practice of the recommendations beyond what would have resulted from just reading the packets alone. This favorable reaction to the training must, of course, be regarded in light of the fact

that the interviewer was known to be a member of the research group that had provided the training.

The minimally trained teachers were asked in the interview to discuss whether or not the training packets alone--the only training medium to which they were exposed--provided enough information to enable them to implement the recommendations. All minimal teachers answered in the affirmative, although approximately half of these teachers added that they would have liked to receive feedback from other teachers and staff or have more specific examples provided. Two teachers said they felt frustrated by not having more help. When minimally trained teachers were asked specifically if they thought additional guidance would have been helpful, two thirds answered affirmatively. These teachers indicated that they would have liked the chance to discuss the recommendations with other teachers and staff, receive feedback, and receive examples and answers for their concerns.

The reactions of maximally and minimally trained teachers to their respective training programs suggests that both groups of teachers saw the opportunity to meet with teachers and staff as an important component of an inservice training program. The need for interteacher communication has been emphasized by many educators concerned with strengthening the teaching profession. Jackson (1971), for example, has noted the plight of teachers as "solitary craftsmen." Lortie (1975) also has described in depth the causes and consequences of the lack of a strong professional culture among teachers. The process and student outcome results of this study do not suggest a clear advantage for the maximal training procedures over the minimal training procedures (Crawford &

Stallings, 1978). Nevertheless, considering the criteria of teachers' personal satisfaction and long-term motivation, the maximal training procedures seem more desirable.

Teachers' Reports on Uncontrolled Factors

The richest and most diverse teacher responses during the interview had to do with uncontrolled factors that might have influenced training implementation. Teachers' descriptions of these factors were in large part unsolicited. The factors can be grouped under the two broad headings of "personal" and "external-organizational." Foremost among "personal" factors seemed to be the teacher's educational philosophy, degree of commitment to practicing the philosophy, and view of the compatibility of the training recommendations with this philosophy. Teachers differed appreciably on these dimensions. Other important personal considerations mentioned frequently had to do with (a) the teacher's past experience, e.g., whether or not the teacher already had been trained in specialized skills, (b) the extent to which the current class situation was familiar and comfortable, (c) outside commitments, such as their families, the recent birth of a child, and various types of community involvement, and (d) such personal problems as poor health or a recent death in the family. While some of the last-mentioned factors may seem far removed from a teacher's professional life, it is clear that they can affect a teacher's response to training.

The "external-organizational" factors received the most emphasis from teachers during the interview. Outstanding, here, was the reference to California's Early Childhood Education (ECE) program. ECE is a comprehensive state-funded reform effort that emphasizes planning that takes place on the school site, staff development, individualized

instruction and record keeping, and parent inclass participation. Over two-thirds of the experimental group teachers (N = 16) mentioned ECE, and approximately two thirds of these teachers (N = 11) indicated that they felt the PTE recommendations were in some way incompatible with the highly individualized and small-group arrangements encouraged by ECE. Some (N = 4) of these teachers said they preferred the more "traditional" PTE approach, while others (N = 2) said they considered ECE a first priority. A few teachers (N = 5) did not consider the two programs to be at odds. Other organizational factors associated with ECE, such as open and multigraded classrooms, were mentioned as potential problems. Teachers indicated that these types of arrangements necessitated (a) a great deal of small group work and (b) the use of team teachers or teacher aides not familiar with the recommendations, thereby contributing to inconsistent implementation. The specialized programs carried out by various school learning centers resulted in another type of instructional confounding. Other external factors, though mentioned rarely, were a high transiency rate (disruptive to continuity) in some classes (N = 2) throughout the school year and parental pressure to carry out certain kinds of instruction (N = 1). While the influence of these factors cannot be precisely measured, it seems plausible that, at least in some cases, they must have imposed limitations on the frequency and consistency with which the training recommendations were implemented. It is possible, for instance, that external-organizational factors were responsible, in part, for the earlier mentioned finding that maximal and minimal teachers were not able to increase significantly their implementation of the more global instructional method recommendations (including some aspects not totally under the control of the teacher--e.g., use of materials, grouping arrangements, etc.) as consistently as the recommendations from the other training areas.

Summary and Conclusions

The usefulness of several kinds of teacher self-report data has been examined. The first set of findings suggested the limited utility of teacher reports about changes in their own behavior when these changes are the focus of a training program. Teachers were aware of the responses that were deemed desirable by experimenters, and therefore they apparently exaggerated their reports of implementation. It also is possible that the teachers themselves were convinced, through their own selective perceptions, that they implemented these behaviors regularly. The remainder of the findings permit the conclusion that many kinds of teacher self-report data are accurate and valuable as supplementary information sources. Teachers seem to be able to describe consistently the characteristics of their general teaching practices and classroom environments despite long intervals of time between measurements. They also are fairly accurate in describing the relative academic standing of their classes.

Teachers' self-reports on their reactions to the training and factors affecting its implementation have undeniable importance. These data help in interpreting results and suggest revisions and considerations in future work. Teacher reports from this study confirm the notion that group interaction is an important motivational component of training programs. The reports also emphasized that, in attempting to introduce change, supervisors deal not only with individual teachers but also with the constraints inherent in the complex organizational structures of the schools (cf. Baldridge & Deal, 1975; Fullan & Pomfret, 1977). When more is known about the implementation patterns of individual teachers, it is possible that the self-report data will show that compliance with training

recommendations was greatest when both individual attitudes and organizational environment were supportive and hospitable.

A final note about self-reporting. The trustworthiness of self-reporting has been a controversial topic for a long time in the social sciences. This controversy is still lively. The best resolution--one which these findings support--is that the trustworthiness of self-reporting depends on just what the self-reporting is about and under what conditions the self-reporting is done. Some recent psychological research, for example, indicates that people usually are unable to report accurately on their higher-order cognitive processing (Nisbett & Wilson, 1977). In contrast, much accumulated evidence in personality research indicates that self-reporting about behavior is at least as reliable and valid as other more common assessment procedures (Mischel, 1968).

Finally, it should be recognized that, when experimenters move from the laboratory into the complex environments of the real world, seeking and recording the self-reports of subjects may offer a relatively efficient way to collect insightful information. As one anthropologist (Den Hollander, 1967) has put it: "We might at least accord [the participants] an equal voice with the foreign observer. It cannot be denied that their knowledge is infinitely greater, and they are, after all, the river that is being mapped" (p. 22).

References

Baldridge, J. V., & Deal, T. E. Overview of change processes in educational organizations. In J. V. Baldridge & T. E. Deal (Eds.), Managing change in educational organizations. Berkeley, Cal.: McCutchan, 1975.

Crawford, J., & Stallings, J. Experimental effects of in-service teacher training derived from process-product correlations in the primary grades. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

Den Hollander, A. J. N. Social description: the problem of reliability and validity. In D. G. Jongmans & D. C. W. Gitkind (Eds.), Anthropologists in the field. Assen, The Netherlands: Van Gorcum and Comp. N. V., 1967.

Fullan, M., & Pomfret, A. Research on curriculum and instruction implementation. Review of Educational Research, 1977, 47, 335-397.

Jackson, P. W. The difference teachers make. In How teachers make a difference. Washington, D.C.: U.S. Government Printing Office, 1971.

Lortie, D.C. Schoolteacher: A sociological study. Chicago: University of Chicago Press, 1975.

Mischel, W. Personality and assessment. New York: John Wiley, 1968.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Stayrook, N., & Crawford, J. An experiment on teacher effectiveness and parent-assisted instruction in the third grade: The observation data. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.